Crystallographic and Biochemical Analysis of a Full Consensus Designed Ankyrin Repeat Protein, of Proteins Involved in Human Host Defence and a Beta-Peptide Hydrolase

DISSERTATION

ZUR
ERLANGUNG DER NATURWISSENSCHAFTLICHEN DOKTORWÜRDE
(DR. SC. NAT.)

VORGELEGT DER
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER
UNIVERSITÄT ZÜRICH

VON
TOBIAS MERZ

VON
BEINWIL AM SEE / LEIMBACH

PROMOTIONSKOMITEE
PROF. DR. MARKUS G. GRÜTTER (VORSITZ)
PROF. DR. RAIMUND DUTZLER
PROF. DR. ANDREAS PLÜCKTHUN
ZÜRICH 2010

INDEX OF CONTENTS

ABBREVIATONS

All abbreviations are either referenced or explained in the text.

SUMMARY / ZUSAMMENFASSUNG

CHAPTER 1

Antibodies belong to the adapted immunity and in general recognize specifically a single antigen. They account for a high detection and clearing efficacy of microbial organisms. Antibodies have become an indispensable tool for biochemical and biophysical applications. However, antibodies or fragments of antibodies often feature unfavourable properties such as susceptibility to reducing agents and peptidases or low thermostability. These bottlenecks have been circumvented by employing other protein scaffolds that serve as alternatives to antibodies; an example – the ankyrin repeat protein – is presented in the first chapter.

KAPITEL 1

Antikörper gehören zur adaptiven Immunität und erkennen normalerweise spezifisch ein einzelnes Antigen. Sie sind verantwortlich für die Erkennung und Eliminierung von Mikroorganismen. Antikörper sind heutzutage unverzichtbare Werkzeuge für biochemische und biophysikalische Anwendungen. Leider haben Antikörper oder Fragmente von Antikörpern nachteilige Eigenschaften wie die Anfälligkeit der Strukturintegrität gegenüber reduzierenden Agenzien, Peptidasen oder hohen Temperaturen. Diesem Umstand versucht man mittels neuartigen Proteingerüsten gerecht zu werden, welche als Alternativen zu Antikörpern Anwendung finden; ein Beispiel – das Ankyrin Molekül – wird im ersten Kapitel präsentiert.

* * * * *

CHAPTER 2

Macrophages, lymphocytes and specialized killer cells (leukocytes) are the innate cellular defence immune system and stimulate the release of signalling molecules upon detection of host danger signals. They release a cocktail of reactive oxygen and nitrogen species as well as peptidases to conquer foreign microorganisms. This cocktail acts non-specifically and is also harmful for the host. Diseases might develop and severe tissue damage can occur if persistent host danger signals cannot efficiently be removed and the peptidase action radius cannot be embanked by endogenous inhibitors. Human neutrophil elastase is such an antimicrobial peptidase that is known to be responsible for a number of lung diseases if not properly controlled. The second chapter provides insight into the development of a potent inhibitor against cathepsin G with a structural and inhibitory study of a redesigned sunflower trypsin inhibitor in complex with porcine pancreatic elastase, human neutrophil elastase and cathepsin G.

KAPITEL 2

Makrophagen, Lymphozyten und spezialisierte Killerzellen (Leukozyten) bilden das zelluläre Immunabwehrsystem und stimulieren die Sezernierung von Botenstoffen nach der Erkennung von potenziell gefährlichen Signalen für den Wirt. Der sezernierte Cocktail aus reaktiven Sauerstoff- und Stickstoffspezies sowie Peptidasen bekämpft fremde Mikroorganismen. Dieser Mix agiert nicht spezifisch, er schädigt auch das Gewebe des Wirts und wirkt solange, bis die Mikroorganismen eliminiert sind. Geschieht dies nicht rasch genug, können sich durch die persistente Aktivität der Peptidasen pathogene Gewebeschäden bilden. Humane neutrophile Elastase ist solch eine antimikrobielle Peptidase, die für verschiedene Lungenkrankheiten verantwortlich ist. Das zweite Kapitel zeigt die Entwicklung eines potenten Inhibitors gegen Cathepsin G mit Resultaten einer strukturellen und enzymatisch-inhibitorischen Studie eines modifizierten Sonnenblumen-Trypsin Inhibitors im Komplex mit Schweinepankreaselastase, humaner neutrophiler Elastase und Cathepsin G.

* * * * *

CHAPTER 3

Chapter three presents the structural investigation of an amino-peptidase from *Sphingosinicella xenopeptidilytica*, named BapA. The peptidase exhibits unusual substrate specificities for beta-peptides. Structures of native BapA, its pro-form and BapA in complex with pefabloc SC, ampicillin and enzymatically hydrolyzed ampicillin provide informative insight into substrate binding. Moreover, a mechanism for the autoproteolytic activation and catalytic substrate hydrolysis is suggested.

KAPITEL 3

Im dritten Kapitel wird die strukturelle Untersuchung einer bakteriellen Aminopeptidase, genannt BapA, von *Sphingosinicella xenopeptidilytica* präsentiert. Die Peptidase hat ein ungewöhnliches Substratspektrum für beta-Peptide. Die Strukturen von nativem BapA, ihrer pro-Form und BapA im Komplex mit pefabloc SC, Ampicillin und enzymatisch hydrolysiertem Ampicillin geben einen profunden Einblick in die Substratbindung. Zudem zeigt ein postulierter Mechanismus die autoproteolytische Aktivierung und katalytische Hydrolyse von Substraten.

* * * * *

CHAPTER 4

A small number of membrane-bound and intracellular multi-domain receptors (Toll, NLRP) that is part of the innate immune system, accounts for the detection of microbial derived structural components. Upon recognition, the receptors assemble into large supra-molecular complexes and provide an activation platform for enzymes of downstream signalling cascades. Death domains are essential domains that enable the

formation of such protein complexes. The fourth chapter focuses on the functional investigation of a homotypic death domain interaction between the pyrin domain of NLRP1 and the adaptor protein ASC.

KAPITEL 4

Eine kleine Anzahl membranständiger und intrazellulärer Rezeptoren (Toll, NLRP) des angeborenen Immunsystems ist für die Erkennung mikrobieller Komponenten zuständig. Nach deren Erkennung bilden diese Rezeptoren supramolekulare Komplexe, die die Aktivierung von Enzymen weiterführender Signalkaskaden steuert. Todesdomänen bilden die wesentlichen Bausteine im Aufbau solcher Komplexe. Das vierte Kapitel widmet sich der funktionellen Untersuchung der homotypen Interaktion zwischen der Pyrin Domäne von NLRP1 und dem Adaptorprotein ASC.

* * * * *

APPENDIX

The appendix section shows preliminary results obtained from ribosome display selection rounds of ankyrin repeat molecules against human caspases.

ANHANG

Im Anhang werden vorläufige Resultate der Selektion von Ankyrin Molekülen mittels Ribosomen Display gegen humane Caspasen gezeigt.

CHAPTER 1

STRUCTURAL ANALYSIS OF A FULL CONSENSUS DESIGNED ANKYRIN REPEAT PROTEIN

INTRODUCTION

Antibodies belong to the adaptive immune system and represent anti-microbial membrane-bound or soluble receptors that are constantly subjected to sequence variability in order to adapt to new polypeptides from hostile invaders. Antibodies are the effector molecules of the humoral immunity. In their soluble form, they are able to bind specifically to foreign microorganisms and proteins, marking them for elimination. In their membrane bound form, they confer antigenic specificity onto B-cells leading to specific proliferation of B-cell clones through the interaction of membrane antibody and antigen. B-cells, upon stimulation, differentiate into long living memory B-cells, enabling a rapid response upon further antigen contact. Specificity, diversity, memory and self and not self recognition are characteristic for adaptive immune responses [1].

Antibodies are indispensable receptor molecules in the adaptive immune system. It seems that their beneficial aspects have also made them indispensable in diverse research fields. Specificity and high affinity towards target proteins are very important properties that allow the investigation of ligands, peptides and proteins and protein complexes. Recent biotechnical developments, however, employ repeat proteins as novel types of binding molecules. Similarly to antibodies, they display highly variable binding surfaces and possess all required features acting as antibody substituents often with improved biophysical properties.

1. From antibodies to designed binding molecules

Antibodies possess a common structure of four peptide chains: two identical polypeptidic light chains of about 25`000 molecular mass and two identical polypeptidic heavy chains of about 50`000 molecular mass (without glycosylations). The chains are covalently linked through disulfide bonds. Each antibody has two antigen binding sites (hypervariable regions). The in sequence variable loops are called complementarity determining regions and are responsible for the recognition of the antigen.

Despite the versatility of antibodies, they frequently possess unfavourable biophysical properties, which limit their application range. A "perfect" antibody should be resistant to aggregation, precipitation and degradation by proteases, is usable in a variety of solvents including detergents and reducing agents, has a long half-life and is easily available and cheap. Antibodies are large and complex molecules that cannot be recombinantly expressed in bacteria due to the presence of disulfide bonds and functional glycosylations [2]. Furthermore, antibody production often requires time-consuming and costly immunizations of animals.

To circumvent whole antibody productions, single polypeptide chain constructs comprising the variable domains (scFv, single chain Fragment variable) of the heavy and light chain of an antibody have been created. These artificial proteins mimic the function of antibodies. Because of their improved biophysical properties they are a popular "antibody format" [2], [3], [4].

Although whole antibodies and scFv are frequently used in standardized protocols as high affinity detection tools, there has recently been a great progress in the development of other domains or scaffolds than immunoglobulins to serve as binding molecules. In 1992, Michaely and Bennett [5] report about ankyrin motifs. "Ankyrin" proteins are structural proteins involved in specific macromolecular recognition. These elongated proteins consist of stacked repeats and mimic in principle antibodies by forming a complementary protein surface well designed for their cognate interaction partner. Because this motif / repeat is found in proteins of the ankyrin family, it was termed ankyrin repeat[1] (Figure 1_1).

Besides ankyrins, another class of repeat proteins is ubiquitously found in nature, the so called leucine rich repeat[2] proteins. Numerous examples of LRR containing proteins are known that interact with various compounds (for reviews see [6], [7]).

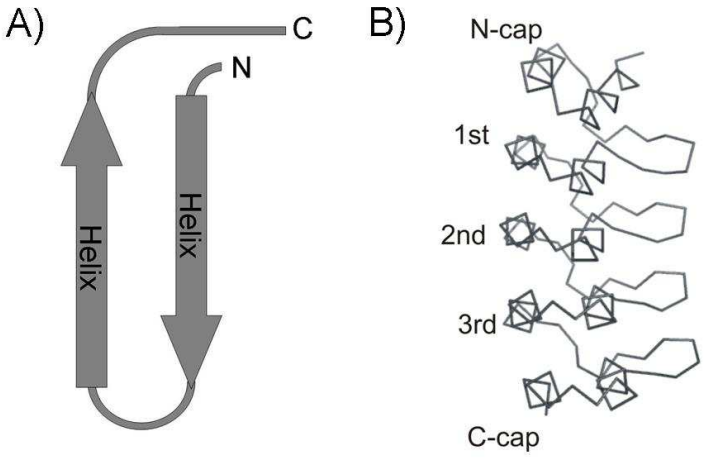


Figure 1_1)
A) Schematic drawing of the secondary structure elements of a single ankyrin repeat; B) The tertiary structure of five stacked ankyrin repeats is shown as ribbon diagram. The figure shows a N3C ankyrin molecule with an amino-terminal *N*-cap and a carboxy-terminal *C*-cap, flanking three internal repeats (ribbon diagram: adapted and modified from [8]).

1.1 Repeat proteins in adaptive immunity?

Few examples of repeat proteins involved in the adaptive immunity are known. Highly diverse LRR proteins, named variable lymphocyte receptors, have been identified in the jawless agnathan sea lamprey and hagfish, which, similarly to antibodies in jawed vertebrates, act in the adaptive immune system [9]. Somatic rearrangements of LRR cassettes within this locus generate unique variable lymphocyte receptor transcripts.

[1] Ankyrin repeat; AR
[2] Leucine rich repeat; LRR

1.2 Repeat proteins as scaffolds for consensus design

Repeat proteins such as ankyrin repeat proteins, tetratricopeptide repeat proteins or leucine rich repeat proteins provide a source of naturally occurring binding molecules that have evolved a successful binding strategy [10]. They have modular extended structures of repeated structural units that stack together to form a continuous target-binding surface. Furthermore, repeat proteins seem to have a high tolerance towards insertion and deletion at the level of entire repeats and insertions at the level of inter-repeat connections [11], [12].

Binding molecules that are based on other protein scaffolds such as the Z domain of staphylococcal protein A (affibodies) [13], anticalins [14] or green fluorescent protein [15] are well expressed and stable but limited by the size of their target-binding surface, reviewed in Skerra et al., 2007 [16].

1.3 Consensus designed ankyrin repeat proteins[3]

The consensus design of ankyrin repeat proteins aims for the generation of combinatorial libraries of polypeptides with highly diversified binding specificities [17], [18].

Ankyrin repeat proteins have ideal folds for consensus design and protein engineering approaches. They are highly abundant in nature with a relatively small number of residues per repeat and a small number or residues per repeat that form the variable surface. This enables a statistical analysis of thousands of sequences. Extensive database searches and structural considerations are implemented into two different approaches for consensus designed ankyrin repeat proteins. Binz et al., 2003, combine a set of diverse repeats with flanking amino- and carboxy-terminal capping repeats from a desoxyribonucleic acid (DNA) binding AR protein GABPβ1 [19], which results in soluble expression [17]. These molecules are known as "NxC" AR proteins, where "N" and "C" refer to the *N*- and *C*-(amino- and carboxy) terminal capping repeats, and "x" to the number of internal ankyrin repeats. The consensus design by Mosavi et al., 2002, a design with identical repeats without caps, yields unfolded proteins in inclusion bodies that require refolding [20]. Structure based substitutions of hydrophobic against charged surface residues significantly improve the solubility of the redesigned ankyrins by Mosavi et al. [21]. These molecules are denoted "nANK", where "n" and "ANK" indicate the number of ankyrin repeats, respectively [20], [21].

1.4 Structure of an ankyrin repeat

The common scaffold of an ankyrin repeat contains 33 amino acids. The secondary structure elements of a single ankyrin module (Figure 1_1) comprise two antiparallel helices, which are connected by a short turn and a linker sequence at the carboxy-terminus connecting the next repeat [22].

All design approaches are based on similar strategies. The lower the variance of an amino acid at a given position and given fold, the more likely it is to be key to fold

[3] Designed ankyrin repeat protein; DARPin

conservation [24]. This is a fundamental principle and a cornerstone for protein design strategies.

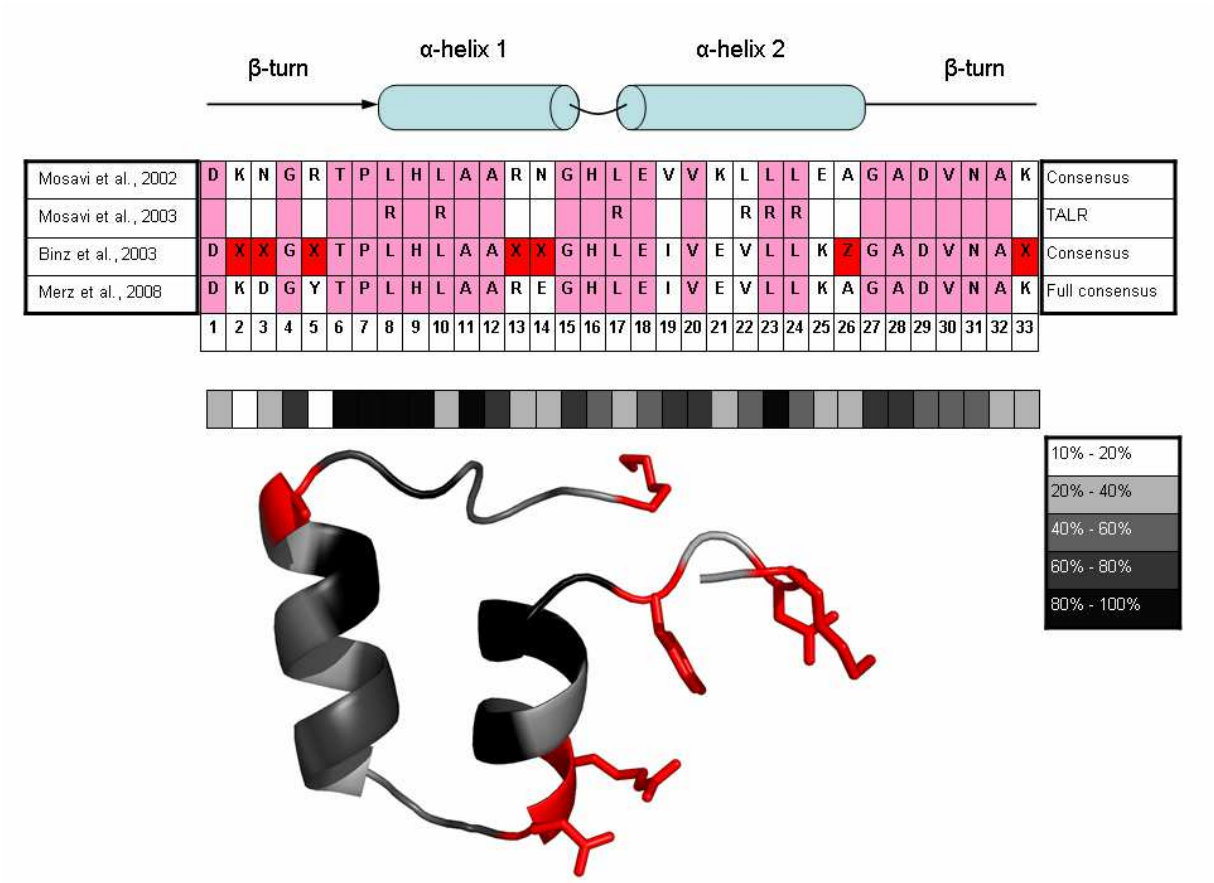


Figure 2_1)
Consensus sequence alignment with structural features of the ankyrin repeat proteins as published by Binz et al., 2003 [17], Mosavi et al., 2002 [20], Mosavi et al., 2003 [21] and Merz et al., 2008 [23]. Various leucines in the *N*-Cap (L8R, L20R and L21R) and in the *C*-Cap (L6R, L15R, L22R) have been replaced by arginines in the TALR sequence [21]. The randomized positions (X), coloured in red, in the consensus sequence designed by Binz et al., 2003 encode for every amino acid except cysteine, glycine and proline; Z within the same amino acid sequence encodes either for histidine asparagine or tyrosine. Pink highlighted amino acids are identical among all sequences. Lower panel: colour coded amino acid conservation throughout the repeat sequence and within the structure.

The choice of amino acids for framework positions, which are necessary to conserve the fold, are chosen upon sequence statistics (relative occurrence) by Mosavi et al., 2002 [20]. Sequence statistics, as well as extensive structural criteria, are taken into account to complement the ankyrin repeat design at less conserved positions with variable "consensus" residues by Binz et al., 2003 [17]. Unsurprisingly, both design strategies result in similar consensus sequences (Figure 2_1). Remaining with the antibody nomenclature, 26 residues constitute the constant part, which is responsible for

framework integrity and seven residues constitute the variable part, displaying the complementary variable surface. The positions of these randomized residues forming the complementary determining region are shown in Figure 3_1. Indispensable prerequisite for all consensus designs is a statistically significant number of available protein sequences. Such consensus designs of repeat proteins are reported for ankyrin repeat proteins [17], [20], [21], tetratricopeptide repeat proteins [25], [26] and leucine rich repeat proteins [27].

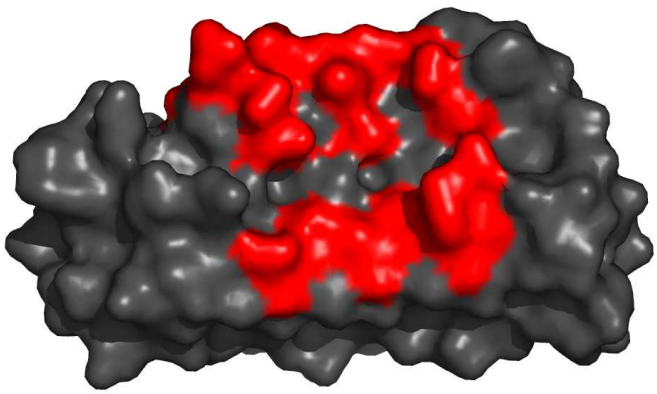


Figure 3_1)
Surface representation of a designed N3C ankyrin molecule (Protein database code 1MJO). The randomized residues pointing towards the concave side of the molecule are coloured in red.

A number of DARPins in complex with their target molecule are known and have been structurally investigated [28], [29], [30]. This shows that *in silico* designed and *in vitro* selected repeat proteins can successfully serve as alternatives to antibodies.

1.5 Consensus versus full consensus

In consensus designed DARPins, variable and even framework positions can differ from one repeat to the next [17]. Since consensus designed DARPins are mostly generated through ribosome display and phage display selection techniques, it is primarily the selection pressure that determines the type of amino acid at a variable position besides additional variability introduced through various polymerase chain reaction (PCR) steps. Hence, one of the selected traits of consensus designed DARPins is their property to bind to target molecules with high specificity and affinity. Affinities in the nanomolar and picomolar range are reported [31], [32].

Full consensus designed DARPins, however, are designed *in silico*. They are not selected against a target protein and are mostly used for biophysical and structural investigations. In full consensus designed DARPins, all internal repeats consist of identical repeat sequences. The molecules are denoted NIxC, in contrast to consensus designed DARPins (denoted as NxC). Here, "N" and "C" also refer to the amino- and carboxy-terminal capping repeats, "I" refers to the identical internal repeats, and "x" indicates the

number of identical internal repeats [23]. Structural considerations, molecular dynamics studies and statistics of relative occurrences from sequence alignments are used to determine the "variable" positions [33]. Such a full consensus designed DARPin is unlikely to have strong binding partner even though this cannot be excluded.

1.6 Chemical and thermal stability

Biochemical analyses show that several consensus and full consensus DARPins display a significantly higher thermostability and chemical stability (guanidinium chloride denaturation) compared to naturally occurring AR proteins [17], [20]. The increase in stability is probably due to the removal of unfavourable interactions (possibly present as a consequence of functional requirements), the removal of irregular insertions or deletions within the repeat and between the repeats, an optimized intra-repeat and inter-repeat packing. Other reasons might be that nature simply has not evolved thus far and that the evolutionary pressure is not stringent enough to efficiently sort out stable from less stable ankyrin molecules. On the other hand, there might be no need to, because naturally occurring ankyrins, unlike the designed ones, have never been selected according to stabilities in systems that are beyond physiological conditions.

The first published structures of designed N3C DARPins (E3_5 and OFF7, a binder against maltose binding protein) emphasize hydrogen bond networks within the β-turn and the conserved TPLH sequence motif (Figure 2_1, positions 6-9) as well as an undisturbed repeat stacking as the major stabilizing features compared to naturally occurring AR proteins [8], [28]. However, E3_19 and NI3C (a full consensus DARPin) also possess these hydrogen bond networks and identical repeat interfaces, but still show significantly different stabilities [34], [35]. Therefore, the composition of the six freely exchangeable surface exposed residues on the concave side of the consensus designed molecule (Figure 2_1) suggests strongest impact on the stability, because all other framework residues are identical. The thermal stability of a full consensus DARPin increases significantly when the variable residues are able to form salt bridge networks [23]. The presence of favourable and unfavourable charge interactions in DARPins was shown in several cases to be a discriminating feature for increased or decreased stability [23], [35].

2. Rationale and aim of the project

The aim of this project is to solve the crystal structure of a full consensus NI3C DARPin and to explain its high thermal stability from a structural point of view.

REFERENCES

1.) Kuby, J., Immunology. fourth edition ed, ed. R.A. Goldsby, Kindt, T.J., Osborne, B.A. . 2000, New York: Plenum Press.

2.) Worn, A. and A. Pluckthun, Stability engineering of antibody single-chain Fv fragments. J Mol Biol, 2001. 305(5): p. 989-1010.

3.) Bird, R.E., et al., Single-chain antigen-binding proteins. Science, 1988. 242(4877): p. 423-6.

4.) Glockshuber, R., et al., A comparison of strategies to stabilize immunoglobulin Fv-fragments. Biochemistry, 1990. 29(6): p. 1362-7.

5.) Michaely, P. and V. Bennett, The ANK repeat: a ubiquitous motif involved in macromolecular recognition. Trends Cell Biol, 1992. 2(5): p. 127-9.

6.) Tschopp, J., F. Martinon, and K. Burns, NALPs: a novel protein family involved in inflammation. Nat Rev Mol Cell Biol, 2003. 4(2): p. 95-104.

7.) Inohara, N. and G. Nunez, NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol, 2003. 3(5): p. 371-82.

8.) Kohl, A., et al., Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci U S A, 2003. 100(4): p. 1700-5.

9.) Pancer, Z., et al., Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature, 2004. 430(6996): p. 174-80.

10.) Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting, Protein repeats: structures, functions, and evolution. J Struct Biol, 2001. 134(2-3): p. 117-31.

11.) Sagermann, M., W.A. Baase, and B.W. Matthews, Structural characterization of an engineered tandem repeat contrasts the importance of context and sequence in protein folding. Proc Natl Acad Sci U S A, 1999. 96(11): p. 6078-83.

12.) Tripp, K.W. and D. Barrick, The tolerance of a modular protein to duplication and deletion of internal repeats. Journal of Molecular Biology, 2004. 344(1): p. 169-78.

13.) Nord, K., et al., Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat Biotechnol, 1997. 15(8): p. 772-7.

14.) Skerra, A., 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. J Biotechnol, 2001. 74(4): p. 257-75.

15.) Abedi, M.R., G. Caponigro, and A. Kamb, Green fluorescent protein as a scaffold for intracellular presentation of peptides. Nucleic Acids Res, 1998. 26(2): p. 623-30.

16.) Skerra, A., Alternative non-antibody scaffolds for molecular recognition. Curr Opin Biotechnol, 2007. 18(4): p. 295-304.

17.) Binz, H.K., et al., Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol, 2003. 332(2): p. 489-503.

18.) Forrer, P., et al., A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Letters, 2003. 539(1-3): p. 2-6.

19.) Batchelor, A.H., et al., The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA.[see comment]. Science, 1998. 279(5353): p. 1037-41.

20.) Mosavi, L.K., D.L. Minor, Jr., and Z.Y. Peng, Consensus-derived structural determinants of the ankyrin repeat motif. Proceedings of the National Academy of Sciences of the United States of America, 2002. 99(25): p. 16029-34.

21.) Mosavi, L.K. and Z.Y. Peng, Structure-based substitutions for increased solubility of a designed protein. Protein Engineering, 2003. 16(10): p. 739-45.

22.) Sedgwick, S.G. and S.J. Smerdon, The ankyrin repeat: a diversity of interactions on a common structural framework. Trends in Biochemical Sciences, 1999. 24(8): p. 311-6.

23.) Merz, T., et al., Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. J Mol Biol, 2008. 376(1): p. 232-40.

24.) Main, E.R., S.E. Jackson, and L. Regan, The folding and design of repeat proteins: reaching a consensus. Curr Opin Struct Biol, 2003. 13(4): p. 482-9.

25.) Main, E.R., et al., Design of stable alpha-helical arrays from an idealized TPR motif. Structure, 2003. 11(5): p. 497-508.

26.) D'Andrea, L.D. and L. Regan, TPR proteins: the versatile helix. Trends Biochem Sci, 2003. 28(12): p. 655-62.

27.) Stumpp, M.T., et al., Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. Journal of Molecular Biology, 2003. 332(2): p. 471-87.

28.) Binz, H.K., et al., High-affinity binders selected from designed ankyrin repeat protein libraries. Nat Biotechnol, 2004. 22(5): p. 575-82.

29.) Schweizer, A., et al., Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism. Structure, 2007. 15(5): p. 625-36.

30.) Sennhauser, G., et al., Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. PLoS Biol, 2007. 5(1): p. e7.

31.) Wyler, E., et al., Inhibition of NF-kappaB activation with designed ankyrin-repeat proteins targeting the ubiquitin-binding/oligomerization domain of NEMO. Protein Sci, 2007. 16(9): p. 2013-22.

32.) Zahnd, C., et al., A designed ankyrin repeat protein evolved to picomolar affinity to Her2. J Mol Biol, 2007. 369(4): p. 1015-28.

33.) Interlandi, G., Wetzel, S. K., Settanni, G., Plückthun, A. and Caflisch, A., Molecular dynamics simulations and chemical denaturation experiments. J Mol Biol, 2008. 375(3): p. 837-54.

34.) Wetzel, S.K., Settanni, G., Kenig, M., Binz, H. K. and Plückthun, A., Folding mechanism of highly stable full consensus ankyrin repeat proteins probed by thermodynamic and kinetic studies. J Mol Biol, 2008. 376(1): p. 241-57.

35.) Binz, H.K., Kohl, A., Plückthun, A., & Grütter, M.G., Crystal structure of a consensus-designed ankyrin repeat protein: Implications for stability. Proteins: Structure, Function, and Bioinformatics, 2006. 65(2): p. 280-4.

doi:10.1016/j.jmb.2007.11.047 J. Mol. Biol. (2008) 376, 232–240

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stabilizing Ionic Interactions in a Full-consensus Ankyrin Repeat Protein

Tobias Merz, Svava K. Wetzel, Susan Firbank, Andreas Plückthun, Markus G. Grütter and Peer R. E. Mittl*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 4 July 2007; received in revised form 28 September 2007; accepted 16 November 2007
Available online
22 November 2007

Full-consensus designed ankyrin repeat proteins (DARPins), in which randomized positions of the previously described DARPin library have been fixed, are characterized. They show exceptionally high thermodynamic stabilities, even when compared to members of consensus DARPin libraries and even more so when compared to naturally occurring ankyrin repeat proteins. We determined the crystal structure of a full-consensus DARPin, containing an N-capping repeat, three identical internal repeats and a C-capping repeat at 2.05 Å resolution, and compared its structure with that of the related DARPin library members E3_5 and E3_19. This structural comparison suggests that primarily salt bridges on the surface, which arrange in a network with almost crystal-like regularity, increase thermostability in the full-consensus NI_3C DARPin to make it resistant to boiling. In the crystal structure, three sulfate ions complement this network. Thermal denaturation experiments in guanidine hydrochloride directly indicate a contribution of sulfate binding to the stability, providing further evidence for the stabilizing effect of surface-exposed electrostatic interactions and regular charge networks. The charged residues at the place of randomized residues in the DARPin libraries were selected based on sequence statistics and suggested that the charge interaction network is a hidden design feature of this protein family. Ankyrins can therefore use design principles from proteins of thermophilic organisms and reach at least similar stabilities.

Keywords: repeat protein; protein stability; salt bridge; thermodynamic stability; X-ray crystallography

Edited by F. Schmid

Introduction

Repeat proteins consist of repeating structural units that stack together to form elongated non-globular domains.[1,2] In contrast to globular proteins, they are not stabilized by interactions between residues that are very distant in sequence; instead, the stabilizing and structure-determining interactions are formed within a repeat and between neighboring repeats. Repeat proteins can be extended in size while still constituting a contiguous domain, making them unique targets for protein engineering. Repeat proteins constitute, next to immunoglobulins, the most abundant natural protein classes specialized in binding.

Because of their abundance and the multiple occurrences of repeats within one protein sequence, a statistical analysis of thousands of sequences can be carried out to design consensus repeats. This has been reported for ankyrin repeat (AR) proteins, tetratricopeptide repeat proteins and leucine-rich repeat proteins.[3–6] The AR is one of the most common protein sequence motifs. This 33-residue motif consists of a β-turn, followed by two antiparallel α-helices and a loop reaching the turn of the next repeat.[7]

*Corresponding author. E-mail address: mittl@bioc.uzh.ch.

Present address: S. Firbank, Institute of Cell and Molecular Biosciences, University of Newcastle, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.

Abbreviations used: DARPin, designed ankyrin repeat protein; AR, ankyrin repeat; PDB, Protein Data Bank; SC, shape complementarity; HB, hydrogen bonds; Gdn·HCl, guanidinium hydrochloride.

Two independent approaches used to apply consensus design strategies to AR proteins have been reported so far.[3,4] Both employ a similar but not identical consensus version of the 33-residue AR.[7] In one approach, large libraries of AR proteins,[3] in which only 26 of the 33 amino acids were specified while 7 were allowed to vary in order to bind to different target molecules, were made.[8]

The potential target interaction residues of a single AR (2, 3, 5, 13, 14 and 33), henceforth named randomized residues, are located on the concave side of designed ankyrin repeat proteins (DARPins).[3] Highly specific binders for a number of different target proteins have been selected using the DARPin libraries[8], and structures of ankyrin–target protein complexes have been determined.[9–12] In this approach, N- and C-capping repeats flanking the randomized repeats that shield the hydrophobic core were employed. Recently, it was shown experimentally that indeed the capping repeats are required for soluble expression in *Escherichia coli*[13] and for shielding of the protein against aggregation.

In another approach, a full consensus was derived from sequence statistics.[4] In these studies, no capping repeats were employed, and the resulting proteins were only soluble at acidic pH.[4] The introduction of positive charges in the C-terminal consensus repeat then allowed the protein to be soluble at neutral pH, but it still had to be produced by *in vitro* refolding from inclusion bodies made in *E. coli*.[5] However, the gain in solubility was accompanied by a significant loss in stability at pH 4.

As described in detail elsewhere in this issue,[14] a full-consensus version of the DARPins, in which randomized residues of the library have now been defined such that every internal repeat has exactly the same sequence, was designed. We denote these proteins as NI_xC, where "N" and "C" refer to the N- and C-terminal capping repeats, respectively, "I" refers to the internal full-consensus repeat and the subscript x gives the number of identical internal consensus repeats. The design of NI_xC proteins and their thermodynamic and kinetic folding properties have been investigated.[14] The N- and C-terminal capping repeats were originally taken[3] from GABP β1 (guanine-adenine binding protein β1) (PDB code 1AWC).[15]

Here we report the crystal structure of the full-consensus DARPin NI_3C at 2.05 Å resolution and compare its features with the structures of two related consensus DARPins E3_5 and E3_19 [Protein Data Bank (PDB) entries 1MJ0 and 2BKG, respectively]. All three molecules were designed using the same framework residues. In contrast to NI_3C, the internal repeats of E3_5[16] and E3_19[17] do not have the same sequences because they contain different residues at the randomized positions. From our structural analysis, we propose that a highly regular array of salt bridges, the overall charge distribution and the binding of sulfate ions significantly contribute to the thermodynamic stability of NI_3C. This study is intended to extend our understanding of stabilizing charge effects in AR proteins.

Results

Overall structure

The final model of the full-consensus NI_3C DARPin structure at 2.05 Å resolution contains all amino acids and side chains of the three identical internal repeats and the flanking N- and C-terminal capping repeats. The capping repeats expose a hydrophilic surface and shield the hydrophobic core of the internal repeats from the solvent to prevent aggregation. As in all AR proteins, the repeats adopt an L-shaped arrangement, which is caused by two almost antiparallel α-helices and a β-turn forming the interrepeat connection (Fig. 1a). Residues positioned at the concave side of the DARPin library (randomized positions) normally mediate specific interactions with selected targets. Based on the statistics of naturally occurring AR protein sequences, they were designed to be highly charged in the full-consensus version.[14] As a consequence, three sulfate ions from the crystallization buffer were found to bind to this site (Fig. 1b).

E3_5/E3_19 consensus *versus* NI_3C full consensus

Sequence differences within repeats between the full-consensus molecule NI_3C and the unselected library members E3_5 and E3_19 occur at positions 2, 3, 5, 13, 14, 26 and 33 (for repeat numbering, see Fig. 2). Residues at these randomized positions depend on the selection process and typically differ from one repeat to the next. In contrast to E3_5 and E3_19, all internal repeats in NI_3C are identical (Fig. 2). As discussed in detail elsewhere in this issue,[14] amino acid types were selected by sequence statistics and structural considerations.

Briefly, at positions 2 and 3, lysine and aspartic acid residues were introduced, respectively. While lysine is the most prevalent amino acid at position 2, aspartic acid is the second most abundant amino acid at position 3 (after asparagine). A negatively charged amino acid was selected for position 3 to balance the positive charges that were introduced at positions 2 and 33. For position 5, which is quite variable, tyrosine was introduced as a spectroscopic probe. In position 13, the second most abundant arginine (after glutamine) was used to compensate for negative charges, and the most abundant glutamic acid was selected for position 14. Alanine was inserted at position 26 because it is most abundant at this position and possesses high helical propensity. For position 33, the most abundant lysine was selected. Figure 2 shows the correspondence between the numbering within a single repeat and the protein sequence.

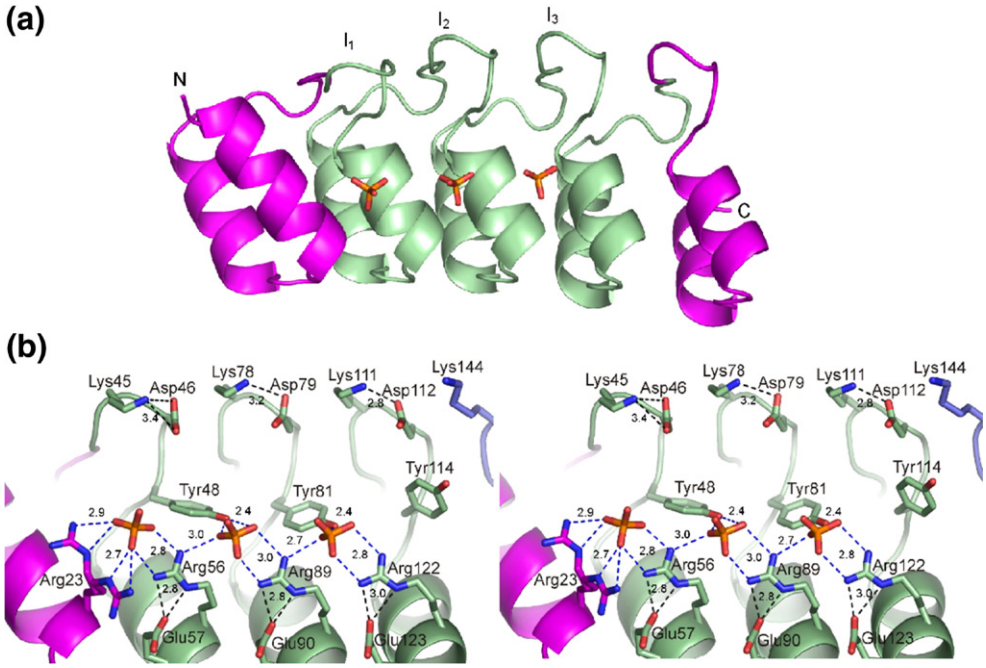


Fig. 1. (a) Ribbon diagram of the NI_3C structure. The terminal and internal repeats are in magenta and green, respectively. (b) Stereo view of the interactions involving the three sulfate ions and the randomized residues. Dashed lines in black denote HB between lysine and aspartate residues at positions 2 and 3, as well as HB between arginine and glutamate residues at positions 13 and 14. Dashed lines in blue denote all interactions of the sulfate ions.

Structural comparison with other AR proteins

The overall structures of NI_3C, E3_5 and E3_19 are very similar. The RMSD values of the pairwise comparisons of Cα atoms between NI_3C/E3_5, NI_3C/E3_19 and E3_5/E3_19 are 0.62 Å, 0.50 Å and 0.60 Å, respectively. Superpositions of the three molecules NI_3C, E3_5 and E3_19 revealed the largest structural differences in the C-terminal capping repeats, which move, in a first approximation, as rigid bodies. This is illustrated by a significant reduction in the overall RMSD values when omitting

```
SEQUENCE          .......20........*.........40..
      NI3C          DLGKKLLEAARAGQDDEVRILMANGADVNAK  :43
N-CAP E3_5          DLGKKLLEAARAGQDDEVRILMANGADVNAT  :43
      E3_19         DLGKKLLEAARAGQDDEVRILMANGADVNAE  :43

POSITION          .........10.........20.........30..
SEQUENCE          ......*.........60.........*......
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAK  :76
1st   E3_5        DNDGYTPLHLAASNGHLEIVEVLLKNGADVNAS  :76
      E3_19       DTYGDTPLHLAARVGHLEIVEVLLKNGADVNAL  :76

POSITION          .........10.........20.........30..
SEQUENCE          ...80.........*.........100.......
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAK  :109
2nd   E3_5        DLTGITPLHLAAATGHLEIVEVLLKHGADVNAY  :109
      E3_19       DFSGSTPLHLAAKRGHLEIVEVLLKYGADVNAD  :109

POSITION          .........10.........20.........30..
SEQUENCE          *.........120.......*.........140
      NI3C        DKDGYTPLHLAAREGHLEIVEVLLKAGADVNAQ  :142
3rd   E3_5        DNDGHTPLHLAAKYGHLEIVEVLLKHGADVNAQ  :142
      E3_19       DTIGSTPLHLAADTGHLEIVEVLLKYGADVNAQ  :142

SEQUENCE          .......*.........160....
      NI3C        DKFGKTAFDISIDNGNEDLAEILQ           :166
C-Cap E3_5        DKFGKTAFDISIDNGNEDLAEILQ           :166
      E3_19       DKFGKTAFDISIDNGNEDLAEILQ           :166
```

Fig. 2. Sequence alignment and numbering of NI_3C, E3_5 and E3_19. The N-terminal capping repeat includes a His_6 tag (MRGSHHHHHHGS; sequence not shown). Identical residues are shown against a black background, whereas randomized residues at positions 2, 3, 5, 13, 14, 26 and 33 in the first, second and third internal repeats are shown against a white background. POSITION: numbering within one repeat; SEQUENCE: numbering throughout the sequence.

the C-terminal repeats. The RMSD values for Cα atoms of residues 1–139 for the pairs NI_3C/E3_5, NI_3C/E3_19 and E3_5/E3_19 are 0.29 Å, 0.39 Å, and 0.36 Å, respectively. Shape complementarity (SC) values were analyzed between internal repeats and fragments thereof (see Materials and Methods). They were, in most cases, higher between internal consensus repeats than between the capping repeats and the adjacent internal consensus repeats (Table 1). This also illustrates the success of consensus design, as the internal repeats can apparently be very well stacked.

Hydrogen bond network

The total numbers of intramolecular hydrogen bonds (HB) in NI_3C, E3_5 and E3_19 are 152, 149 and 152, respectively, excluding those involved in binding sulfate ions. Interrepeat HB are least frequent in NI_3C, with 14 HB, compared to 16 and 17 HB in E3_5 and E3_19, respectively. Due to the high sequence identity, the hydrogen-bonding pattern is very similar in all three structures, as indicated by 119 common HB. The main-chain hydrogen-bonding pattern of the conserved TPLH sequence motif (residues 6–9) at the beginning of the first α-helix of every repeat is identical in NI_3C, E3_5 and E3_19. Small structural differences between different types of amino acids at the randomized positions enable the formation of specific HB. In NI_3C, E3_5 and E3_19, there are 22, 17 and 19 specific HB, respectively.

A detailed analysis of the total number of HB at the randomized positions reveals that 10 out of 34 HB in NI_3C, 5 out of 30 HB in E3_5 and 6 out of 30 HB in E3_19 are formed by side chains, again excluding those involved in binding sulfate ions. In contrast, the majority of the randomized residues contribute with their backbone atoms to the overall hydrogen-bonding network.

The rather modest number of side-chain HB from randomized residues in E3_5 and E3_19 shows that these residues are more relevant to the binding of the target molecules rather than to the formation of extended hydrogen-bonding networks on the surface of the molecules (Table 1). The side chains at the previously randomized positions specified in the full-consensus NI_3C molecule are involved in highly regular charge–charge interactions. This charge network is extended by sulfate ions from the crystallization buffer by bridging four arginine and two tyrosines residues in the first two repeats (Fig. 1b).

Charge network

Previous reports on DARPins[3,4,16] emphasized the highly regular hydrogen-bonding patterns between "framework" parts (i.e., contributed by the constant part of the sequence as important for stability). The designed ankyrins displayed a significantly higher thermodynamic stability compared to naturally occurring AR proteins,[3,4] which do not have such regular hydrogen-bonding networks and show greater variability.

The full-consensus AR NI_3C possesses, in addition, a highly regular charge distribution, which spans about half of the protein surface and causes an even higher thermodynamic stability compared to other library members. NI_3C contains a regular array of lysine and aspartic acid residues at positions 2 and 3 in the β-turns of internal repeats, as well as arginine and glutamic acid residues at positions 13 and 14 at the C-terminal ends of helices in the concave binding region. Arg23 positioned in the N-terminal capping repeat and Lys144 in the β-turn of the C-terminal capping repeat extend the charge network. The electron density of Arg23 suggests a double conformation. Both conformations allow the formation of salt bridges with a first sulfate ion. In the concave binding region, three equally spaced sulfate ions form salt bridges with Arg56, Arg89 and Arg122 (corresponding to position 13 of the internal repeat) and HB with Tyr48 and Tyr81 (position 5 in the internal repeat). In addition, these arginine residues participate in salt bridges with Glu57, Glu90 and Glu123 (corresponding to position 14 of the internal repeats) (Fig. 1b).

Thermal denaturation measurements

The thermal denaturation of NI_3C was monitored at 222 nm using CD spectroscopy. Under physiological conditions, NI_3C, unlike E3_5 and E3_19,[3,17] did not show a transition in thermal denaturation experiments.[14] In order to denature the protein

Table 1. Summary of the structural analysis between NI_3C, E3_5 and E3_19

Property	Ankyrins	NI_3C	E3_5	E3_19	
Identity (%)	NI_3C	100	89	88	
	E3_5	88	100	88	
	E3_19	87	87	100	
SC (shape complementary)	NI_3C	0.714 (SC_{N-1})	0.839 (SC_{1-2})	0.803 (SC_{2-3})	0.705 (SC_{3-C})
	E3_5	0.722 (SC_{N-1})	0.762 (SC_{1-2})	0.805 (SC_{2-3})	0.798 (SC_{3-C})
	E3_19	0.702 (SC_{N-1})	0.812 (SC_{1-2})	0.769 (SC_{2-3})	0.739 (SC_{3-C})
HB (overall)	Total	152	149	152	
	Unique	22	17	19	
	Common				119
HB (randomized residues)	Total	34	30	30	
	Side chain	10	5	6	
	Main chain	24	25	24	

completely and to reach the posttransition baseline, thermal denaturation measurements were performed in the presence of 4 M guanidinium hydrochloride Gdn·HCl. Denaturation experiments of NI_3C in Gdn·HCl had revealed a complex unfolding mechanism, which was interpreted as a partial unfolding of the C-terminal capping repeat prior to the main transition.[13,14]

A remarkable shift towards higher transition temperatures in the thermal melting curves of NI_3C was found with increasing sulfate concentrations (Fig. 3a and b). The increase in melting temperature is much more modest with increasing sodium chloride concentrations. The analysis of T_m as a function of ionic strength revealed that T_m increases linearly with ionic strength, which is in agreement with previous studies,[18] but the slope of the regression line is significantly higher for sulfate (0.0092 °C mM^{-1}; $R^2=0.9884$) than for phosphate (0.0064 °C mM^{-1}; $R^2=0.9767$) or chloride (0.0021 °C mM^{-1}; $R^2=0.9268$) ions (Fig. 3c). In contrast, E3_19, which does not have specific sulfate-binding sites and whose limited stability has been proposed to be a consequence of local repulsions,[17] is also stabilized by sodium chloride and sulfate (Fig. 3d). In this case, the melting point depends only on ionic strength and is independent of salt type.

Discussion

Thermal and chemical stabilities are highly relevant to the biotechnological and biomedical applications of proteins and are thus an important design goal. Consensus design exploits sequence statistics and is based on the assumption that the most frequently occurring residues in a first approximation are correlated with molecules of high thermal stability.[19] As most random mutations are destabilizing, only those protein variants that have stabilizing residues elsewhere can tolerate them. The consensus of all naturally occurring sequences would be expected to reflect these favorable residue combinations. A structural inspection is still required to avoid designing mutually exclusive residues. Compared to naturally occurring ankyrin proteins with similar numbers of repeats, consensus-designed AR proteins showed higher thermodynamic stabilities and increased stabilities towards chemical denaturants.[5,16]

The previous consensus design of DARPins has been limited to "framework" positions. The original goal was to create a library with the potential to recognize a wide range of target molecules. Consequently, residues at the randomized positions were allowed to vary and exerted attractive and repulsive charge interactions. Nonetheless, key residues specifying intrarepeat and interrepeat HB were present in every one of the internal DARPin repeats, but not necessarily in natural AR proteins, thereby partially explaining the greater stability of the designed library molecules.[3,16]

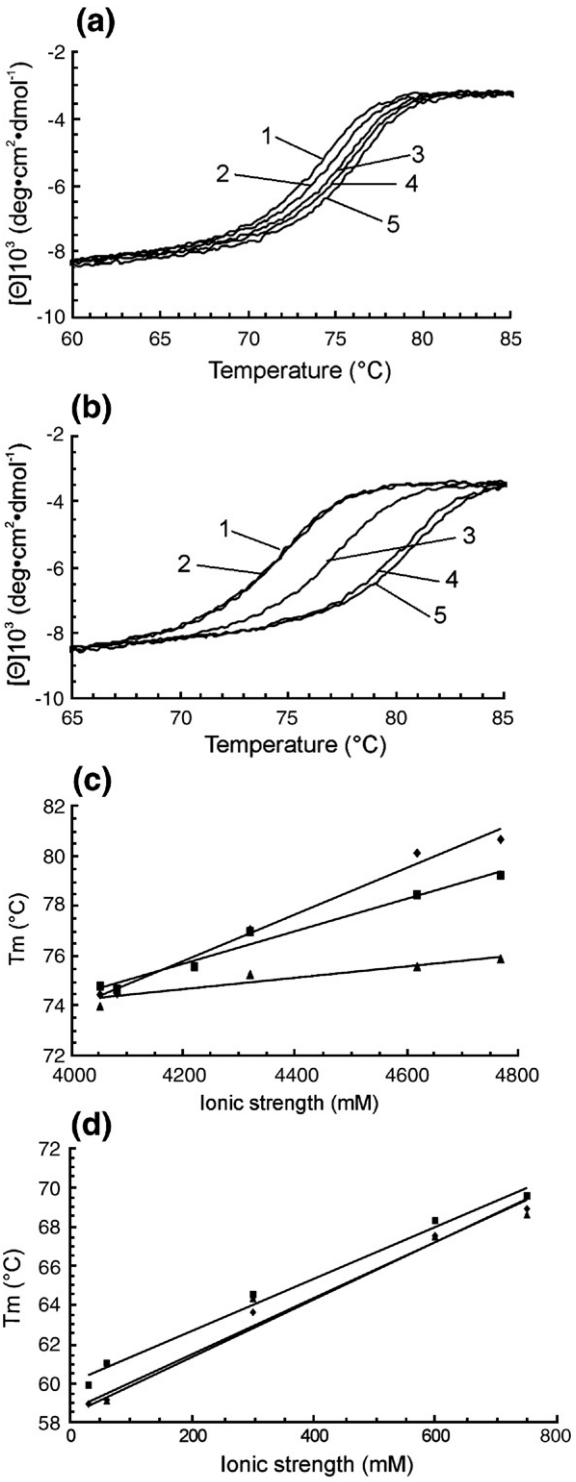


Fig. 3. Thermal melting curves of NI_3C in 20 mM Hepes, 4 M Gdn·HCl (pH 7.4) and different salt conditions: (a) sodium chloride: 30 mM (curve 1), 60 mM (curve 2), 300 mM (curve 3), 600 mM (curve 4) and 750 mM (curve 5); (b) sodium sulfate: 10 mM (curve 1), 20 mM (curve 2), 100 mM (curve 3), 200 mM (curve 4) and 250 mM (curve 5). (c) Melting points (T_m) of NI_3C plotted as a function of ionic strength for different sodium chloride (▲), sodium sulfate (♦) and sodium phosphate (■) concentrations. (d) Melting points of E3_19 as a function of ionic strengths in the same buffers as for NI_3C, but omitting Gdn·HCl. The estimated errors on T_m are ±0.5 °C.

In the present work, we investigated the structural consequences of a full-consensus design where the sequence variability of internal repeats has been eliminated. In order to correlate structural features with the increased thermal stability of the full-consensus protein, we compared the crystal structures of the full-consensus ankyrin NI_3C with the previously published consensus ankyrins E3_5[16] and E3_19[17] (PDB codes 1MJO and 2BKG, respectively). A comparison of RMSD values (Table 1) and visual inspection of Cα–Cα distance plots (data not shown) indicate no major structural differences between the molecules, except within the last repeat. They have almost the same number of HB, and most of them are common between all three molecules. Contacts between repeat interfaces were measured qualitatively by analyzing shape complementarities.[20] Differences in SC values between internal repeats of NI_3C, E3_5 and E3_19 indicate that the stackings of internal repeats differ due to minor side-chain rearrangements. It is unlikely, however, that these differences explain the higher stability of NI_3C.

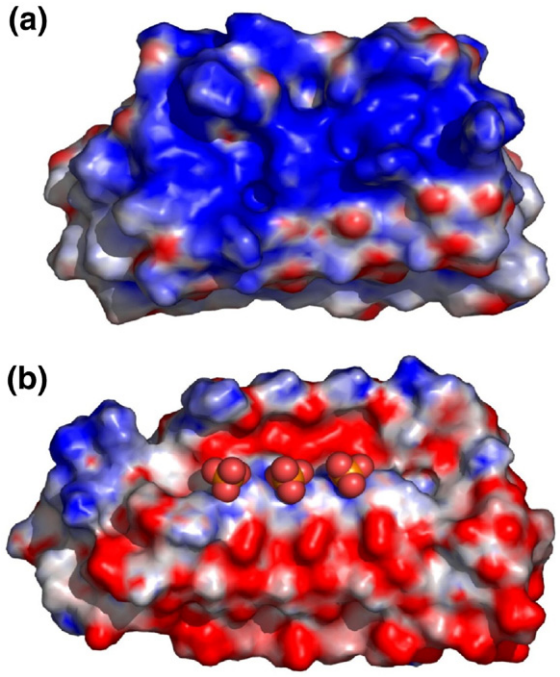


Fig. 4. Electrostatic surface potentials, calculated using the Adaptive Poisson–Boltzmann Solver[21] of (a) 4ANK (PDB code 1N0R) and (b) NI_3C. Both representations show the concave binding site with the N- and C-termini on the right and left sides, respectively. Bound sulfate ions in the NI_3C molecule are shown as spheres. Both surfaces were in blue and red for positive and negative electrostatic charge densities (same scale), respectively.

Charges and charge networks

The major difference between NI_3C, E3_5 and E3_19 is observed in the surface charge distribution. As explained above, the NI_3C design was based on sequence statistics, leading to a charge network that must have been selected during the evolution of AR proteins. One may speculate that residues involved in stabilizing charge networks are most strongly selected and thus dominate sequence statistics of surface residues, whereas residues that are required for target recognition cancel out in the alignments of whole sequence families.

In the work of Mosavi *et al.*[4], a different full-consensus sequence had been derived. In their work, positions 3 and 14 were chosen as uncharged (both asparagine), and an additional charge was introduced by an arginine at position 5 (tyrosine in this work). Additionally, there were two charge reversals in the framework of helix 2, which are lysine at position 21 (here glutamic acid because of its occurrence in GA-binding protein β1) and glutamic acid at position 25 (here lysine because of the negative charge at position 21). In the design of Mosavi *et al.*[4] (PDB entries 1N0Q and 1N0R), these residues generated a cluster of positively charged residues on the concave side of the molecule that might be electrostatically unfavorable (Fig. 4a). In contrast to this, the surface potential of the NI_3C molecule is more balanced in this area (Fig. 4b), especially when considering the protein without the sulfate ions.

Nature has used a multitude of strategies for the adaptation of proteins to life at high temperatures.[22] Surface charges were, for quite some time, considered as rather unimportant for protein stability. It was argued that the high dielectric constant of the solvent would decrease the strength of charge–charge interactions.[23] Recently, it was shown that computational redesign of surface charges can improve protein stability significantly,[24] and structural genomics data on proteins from *Thermotoga maritima* revealed a significant increase in the density of salt bridges in proteins from this organism compared to their mesophilic counterparts.[25] In contrast, the importance of oligomerization order, HB and secondary structure was found to be less pronounced than previously assumed. Similar results were found in another comprehensive structural bioinformatics study[26] and in more specialized studies on individual protein families.[27–30] The high number of salt bridges in hyperthermophilic proteins was explained by the diminishing desolvation penalty for salt bridges at increasing temperatures.[31] Therefore, the contribution of salt bridges to protein stability becomes more important at higher temperatures. NI_3C has the smallest and most balanced negative charge excess among all three molecules studied here and, thus, essentially fewer local repulsions. A higher charge imbalance on E3_19 compared to E3_5 was proposed to be the major cause of its decreased stability.[17,32]

Another possible explanation for the extraordinarily high temperature stability, even in the absence of sulfate ions, is the flexibility of the salt bridge network, giving it an entropic advantage. Although the surface-exposed ion pair network looks rigid in

the present NI_3C structure, only modest side-chain rearrangements would be necessary to swap between intrarepeat and interrepeat salt bridges. We suggest that the sum of all favorable charge interactions (and the absence of unfavorable ones) of the full-consensus ankyrin NI_3C is responsible for its significantly increased thermal stability. It is remarkable that the sequence statistics, at least in part, mirror these favorable properties.

Sulfates in the charge network

Three regularly arranged sulfate ions form the core of the salt bridge network on the concave side of the molecule. Thermal melting experiments in the presence and in the absence of sulfate ions clearly confirm the contribution of sulfate to the thermal stability of the full-consensus NI_3C ankyrin. It should be noted that thermal stability is already very high in the absence of sulfate. Even though the surface potential is balanced (Fig. 4b), the protein can bind three sulfate ions, which increase the total negative charge excess of the molecule from 12 e^- to 18 e^-. Because of the alternating arrangement of the charges, sulfate binding is energetically favored, and the stabilizing effect may be rationalized by the very short hydrogen-bonding distance of 2.4 Å between the sulfate ions and tyrosine residues at position 5 (Fig. 1b).

Summary and conclusion

The crystal structure and the thermal denaturation studies of the designed full-consensus AR protein NI_3C suggest that the extended salt bridge network positively influences the stability of the protein. This protein can only be denatured by heating in the presence of 4 M Gdn·HCl. NI_3C was designed based mainly on sequence statistics, suggesting that this charge network among previously randomized residues must also be a hidden design feature of natural ARs. In naturally occurring AR proteins, this feature is no longer visible because the sequences have drifted due to functional selection. Designing such charge interaction networks may thus be a strategy for the development of extremely stable proteins for biomedical and other biotechnical applications. However, since many of these residues normally participate in binding, a compromise between function and thermal stability by surface charge interactions will have to be found.

Materials and Methods

Expression and purification

The full-consensus NI_3C ankyrin was expressed in *E. coli* strain XL-1 blue in LB medium supplemented with 1% glucose, 100 μg/ml ampicillin and 15 μg/ml tetracycline. After growing to an OD_{600} of approximately 0.8–1.0 at 37 °C, the cultures were induced with 0.5 mM IPTG, and growth continued for 5 h. After centrifugation and resuspension in 50 mM Tris, 500 mM NaCl and 20 mM imidazole at pH 8.0, the cells were disrupted using a French press. The protein was purified in self-packed columns containing a Ni-NTA matrix (Qiagen) in accordance with the manufacturer's instructions. The protein was further purified by gel filtration on a HiLoad 26/60 Superdex 75 (Pharmacia Biotech) with an ÄKTA Prime system (Amersham, Pharmacia Biotech) in 10 mM Hepes and 10 mM NaCl (pH 7.4) and concentrated to 20 mg/ml using a Centricon (Millipore USA) with 3-kDa molecular mass cutoff.

Crystallization

Crystals for X-ray diffraction data collection were grown in 24-well crystallization plates using the hanging-drop vapor-diffusion method. A 1-μl protein solution was mixed with 1 μl of reservoir solution containing 2.7 M $(NH_4)_2SO_4$ and 100 mM Tris (pH 8.5). The crystals grown under these conditions were soaked in reservoir solution supplemented with 20% glycerol as cryoprotectant and flash-cooled in a nitrogen stream at 100 K. The X-ray diffraction data of one single crystal were collected using CuKα radiation generated by a Nonius FR591 rotating anode generator.

Structure solution and refinement

Diffraction data up to 2.05 Å resolution on a total of 180 frames were recorded within an oscillation range of 1°. The data were processed with program XDS.[33] A Matthews coefficient of $V_M = 2.4$ Å^3 Da^{-1} suggested one molecule in the asymmetric unit. Initial phases were obtained by molecular replacement using the program

Table 2. Statistics for data collection and refinement

Data collection	
Space group	$P6_1$
Cell dimensions	$a = 74.48$ Å, $b = 74.48$ Å, $c = 50.99$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$
Resolution limits (Å)	64.5–2.05
Observed reflections	78,634 (overall); 10,232 (unique)
Completeness (%)	99.8 (98.2)
Redundancy	7.7 (7.1)
R_{sym} (% on I)	7.3 (37.0)
Refinement	
Resolution range (Å)	64.5–2.05
R-factor/R_{free} (%)	18.6/22.6
Ordered water molecules	110
RMSD from ideal geometry	
Bond length (Å)	0.010
Bond angles (°)	1.109
Average B-factor (Å^2)	26.3
Residues in the most favored region (%)	91.9
Residues in the additionally allowed region (%)	8.1
Residues in the generously allowed region (%)	0.0
Residues in the disallowed region (%)	0.0

Numbers inside parentheses refer to the highest resolution shell (2.1–2.05 Å).

AMoRe,[34] with the structure of E3_5 (PDB code 1MJO) as a search model. The structure was refined to final *R*-factors of R_{work}=18.6% and R_{free}=22.6% in the space group $P6_1$ using the program REFMAC5.[35] Model building was performed with program O.[36] Statistics for diffraction data and structure refinement are summarized in Table 2. The program PROCHECK[37] was used to evaluate deviations from standard geometries, and the programs HBPLUS[38] and SC[20] were used to analyze the hydrogen-bonding pattern and calculations of surface complementarities. Although the SC program was originally developed to calculate the surface complementarity between separated subunits, we used this method to investigate intramolecular shape complementarities. For this purpose, the molecule was split into fragments by inserting artificial chain breaks between residues 31/32, 64/65, 97/98 and 130/131. The SC_{N-1}, SC_{1-2}, SC_{2-3} and SC_{3-C} values refer to the interface SC values of neighboring repeats. Figures were generated with PyMOL.[39]

Thermal denaturation experiments

Thermal denaturation experiments were performed with a Jasco J-715 instrument (Jasco, Japan). CD data were recorded at a protein concentration of 40 μM in the temperature range between 5 °C and 95 °C within 90 min at a wavelength of 222 nm (measuring interval, 10 s; bandwidth, 2 nm). The buffers contained 20 mM Hepes, 4 M Gdn·HCl (pH 7.4) and salt concentrations of between 30 and 750 mM sodium chloride or of between 10 and 250 mM sodium sulfate. The same buffers without Gdn·HCl were used for heat denaturation experiments with E3_19. Ionic strengths were calculated as follows: $I=1/2\sum_i c_i z_i^2$, where c=concentration and z=charge of ion.

Protein Data Bank accession number

Coordinates and diffraction data have been deposited at the PDB under accession number 2QYJ.

Acknowledgement

Financial support from the Swiss National Science Foundation and the Baugartenstiftung (Zurich, Switzerland) is gratefully acknowledged.

References

1. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131.
2. Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins: Struct. Funct. Genet.* **17**, 363–374.
3. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503.
4. Mosavi, L. K., Minor, D. L., Jr & Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
5. Mosavi, L. K. & Peng, Z. Y. (2003). Structure-based substitutions for increased solubility of a designed protein. *Protein Eng.* **16**, 739–745.
6. Stumpp, M. T., Forrer, P., Binz, H. K. & Plückthun, A. (2003). Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* **332**, 471–487.
7. Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316.
8. Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grütter, M. G. *et al.* (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J. Biol. Chem.* **280**, 24715–24722.
9. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P. *et al.* (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582.
10. Kohl, A., Amstutz, P., Parizek, P., Binz, H. K., Briand, C., Capitani, G. *et al.* (2005). Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure*, **13**(8), 1131–1141.
11. Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E. *et al.* (2007). Inhibition of Caspase-2 by a designed ankyrin repeat protein: specificity, structure and inhibition mechanism. *Structure*, **15**(15), 625–635.
12. Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M. G. (2007). Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol.* **5**(1), e7.
13. Interlandi, G., Wetzel, S.K., Settanni, G., Plückthun, A., Caflisch, A. (2007). Molecular dynamics simulations and chemical denaturation experiments. *J. Mol. Biol.* doi:10.1016/j.jmb.2007.09.042.
14. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. (2008). Folding and unfolding mechanism of highly stable full consensus ankyrin repeat proteins. *J. Mol. Biol.* **376**, 241–257.
15. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L. & Wolberger, C. (1998). The structure of GABPalpha/beta: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science*, **279**, 1037–1041.
16. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl Acad. Sci. USA*, **100**, 1700–1705.
17. Binz, H. K., Kohl, A., Plückthun, A. & Grütter, M. G. (2006). Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. *Proteins: Struct. Funct. Genet.* **65**, 280–284.
18. Tadeo, X., Pons, M. & Millet, O. (2007). Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation. *Biochemistry*, **46**, 917–923.
19. Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A. (2004). Consensus design of repeat proteins. *ChemBioChem*, **5**, 183–189.
20. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.

21. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
22. Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738–748.
23. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science*, **310**, 638–642.
24. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T. *et al.* (2006). Protein stability and surface electrostatics: a charged relationship. *Biochemistry*, **45**, 2761–2766.
25. Robinson-Rechavi, M., Alibes, A. & Godzik, A. (2006). Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermotoga maritima*. *J. Mol. Biol.* **356**, 547–557.
26. Alsop, E., Silver, M. & Livesay, D. R. (2003). Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis. *Protein Eng.* **16**, 871–874.
27. Cheung, Y. Y., Lam, S. Y., Chu, W. K., Allen, M. D., Bycroft, M. & Wong, K. B. (2005). Crystal structure of a hyperthermophilic archaeal acylphosphatase from *Pyrococcus horikoshii*—structural insights into enzymatic catalysis, thermostability, and dimerization. *Biochemistry*, **44**, 4601–4611.
28. Corazza, A., Rosano, C., Pagano, K., Alverdi, V., Esposito, G., Capanni, C. *et al.* (2006). Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile *Sulfolobus solfataricus*. *Proteins: Struct. Funct. Genet.* **62**, 64–79.
29. Karlstrom, M., Steen, I. H., Madern, D., Fedoy, A. E., Birkeland, N. K. & Ladenstein, R. (2006). The crystal structure of a hyperthermostable subfamily II isocitrate dehydrogenase from *Thermotoga maritima*. *FEBS J.* **273**, 2851–2868.
30. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P. C. (1998). Protein thermostability in extremophiles. *Biochimie*, **80**, 933–941.
31. Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489–502.
32. Yu, H., Kohl, A., Binz, H. K., Plückthun, A., Grütter, M. G. & van Gunsteren, W. F. (2006). Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. *Proteins: Struct. Funct. Genet.* **65**, 285–295.
33. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
34. Navaza, J., Panepucci, E. H. & Martin, C. (1998). On the use of strong Patterson function signals in many-body molecular replacement. *Acta Crystallogr. Sect. D*, **54**, 817–821.
35. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D*, **55**, 247–255.
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A*, **47**, 110–119.
37. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
38. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
39. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA, USA.

CHAPTER 2

INNATE ANTIINFLAMMATORY RESPONSES

INTRODUCTION

Physical and simple physiological barriers are an organisms` first line of defence. The epidermis (outer layer) and dermis (inner layer) provide mechanical barriers that prevent entry of microbes and retard bacterial growth by keeping their environment slightly acidic (pH 3-5). Mucus membranes entrap foreign microorganisms hampering them to attach to the underlying epithelial layer, and the normal flora competes with microbes for attachment sites as well as nutrients. Cilia in the lungs actively propel microorganisms out of the body. Physiologic barriers such as high body temperature, low pH in the stomach, digestive enzymes and chemical mediators like interferons and complement inhibit or prevent growth of several microorganisms. These relatively simple non-specific factors and mechanisms are constitutively acting against foreign microorganisms even in the absence of hostile stimuli or from signals that result from an anti-inflammatory response [1].

Further non-specific host defence lines include circulating phagocytic cells, inflammatory barriers as well as broadly acting proteolytic enzymes that are released by neutrophils. These mechanisms also belong to the first line of defence although they use a specific set of signalling molecules and target specific receptors.

1. The cellular non-specific host response

The most abundant circulating granulocyte cell type of the innate immune system are neutrophils (90%), besides eosinophils (2-5%) and basophile granulocytes (less than 0.2%) [2]. Neutrophils and sometimes eosinophils are recruited from the circulation to the site of infection as a first line of defence against foreign invaders [3], [4], [5]. Their primary defence mechanism is the release of nitrogen and oxygen radicals, potent antimicrobial compounds and peptidases. They also remove pathogens by phagocytosis. Neutrophils secrete a variety of cytokines, peptidases and chemokines that alter the connective tissue for the recruitment of inflammatory and immune cells [4], [6]. Although neutrophils are powerful and crucial effector cells in the host defence against infection, they also have the potential to damage host tissues and cells, since molecules released to thwart invaders cannot differentiate self from foreign.

Besides neutrophils, different cell types including macrophages, mast cells, basophiles and eosinophiles secrete a variety of cell-type specific mediators and signalling molecules, which activate resting cells, lead to the production of other signalling molecules or to the expression of surface receptor molecules. Colony stimulating factors released by macrophages[1] do not only act on progenitor cells, and cause proliferation and differentiation, but additionally act on mature cells, promoting phagocytosis, enzyme production and neutrophil adhesion [7], [8]. Neutrophil adhesion and O^{2-} release is also initiated by endothelial cells [9]. On the other side, tumor necrosis

[1] Granulocyte macrophage colony stimulating factor; GM-CSF

factor alpha[2] released by neutrophils can activate and stimulate endothelial cells to express adhesion molecules [10] or to induce apoptosis [11]. Mast cells accumulate a broad array of mediators and signalling molecules, in particular histamine. Various peptidases and TNF-α, released by mast cells cause the major signs of inflammation. Similarly to mast cells, activated eosinophils and basophiles release a variety of cell specific cytotoxic molecules such as major basic protein, eosinophil peroxidase, and proinflammatory cytokines (IL-3[3], IL-4, IL-6, IL-16, TNF-α, GM-CSF) and monocyte chemoattractant protein 1 [12], [13], [14] . Basophile restricted IL-4 and IL-13 stimulate the differentiation of Th_2 cells and IgE[4] synthesis [15].

1.1 Polymorph nuclear neutrophils[5]

The lifespan of circulating non-stimulated neutrophils is rather short. They have a 14 day development phase in the bone marrow followed by a stay in a temporary storage pool before getting released into the blood. In the absence of any bacterial infection they return to the bone marrow or enter the liver after 12-14 hours to undergo apoptosis [16], [17]. The shrinkage into apoptotic bodies and the subsequent phagocytosis by local macrophages prevents the onset of tissue damage by lytic factors [18].

The pool of circulating non-stimulated neutrophils can differentiate into resting polymorph nuclear neutrophils upon recognition of a priming stimulus and requires subsequent stimuli to become active PMN. Three different states of PMN are known: the resting, primed and activated state. The priming process has been demonstrated by *in vitro* administering PMN with sub-stimulatory concentrations of pharmacological agents allowing the cells to respond to a second stimulus. One characteristic feature of PMN is the release of reactive oxygen and nitrogen species, called the oxidative burst. At the very end of the priming events, it is executed upon formyl-methionyl-leucyl-phenylalanine (fMLP) exposure after a pre-treatment of a number of inflammatory mediators such as TNF-α, GM-CSF, lipopolysaccharides or a low dose of fMLP [19], [20]. Formylated peptides released by foreign organisms or by opsonins are recognized by PMN surface receptors [20]. Above a certain threshold of bound peptides the primed state is induced by a rapid increase of cytoplasmic Ca^{2+} concentrations and the simultaneous decrease of the cytoplasmic pH [21], [22], [23]. Interestingly, fMLP induced PMN move up the chemoattractant gradient towards higher concentrations, whereas GM-CSF induced cells start to migrate randomly, showing that bacterial derived stimuli (fMLP) generate a stronger and more directed response [24].

1.2 The oxidative burst

The transition of neutrophils from a resting state in the circulation to an activated state at the site of infection is triggered by an ordered sequence of signals of circulating priming

[2] Tumor necrosis factor alpha; TNF-α
[3] Interleukin; IL
[4] Immunoglobulin; Ig
[5] Polymorph nuclear neutrophil; PMN

and extravascular stimuli. This prevents premature degranulation and collateral host tissue damage upon release of the non-specific arsenal of antimicrobial weapons such as reactive oxygen and nitrogen species, cytotoxic and proteolytic enzymes.

IL-18, a member of the IL-1 family, is involved in the induction of the oxidative burst [25] and to be an important neutrophil mediator in inflammatory diseases [26]. It is concomitantly released with IL-1β and other cytokines such as GM-CSF and TNF-α from endothelial or epithelial cells that sense bacterial antigens for example by Toll-like receptors or intracellular pathogen associated molecular pattern detection platforms such as the inflammasomes (see chapter 4). Besides the importance of soluble factors that trigger the release of reactive oxygen and nitrogen species, PMN only release their destroying cargo at the region of maximal chemoattractant concentration in a primed state. This strongly implies the involvement of membrane bound receptors, which themselves trigger the initiation of degranulation upon contact formation [4].

The nicotinamide adenine dinucleotide phosphate (NADPH) oxidase catalyzes chemical reactions of the oxidative burst. It is a multimeric protein complex that assembles on the membrane after cell priming. The importance and proper functionality of the NADPH oxidase is exemplified by its dysfunction causing chronic granulomatous disease (recurrent infections due to incapability of phagocytes to kill invading microbes) [27].

The phagocyte NADPH oxidase generates superoxide anions using molecular oxygen and NADPH according to the following reaction

$$NADPH + 2\,O_2 \rightarrow NADP^+ + 2\,{}^{\bullet}O_{2^-} + H^+$$

To prevent host tissue damage by the release of the highly reactive oxygen radicals, cells, that are capable of phagocytosis, "kill" the engulfed bacterium within protected areas. The lytic and killing process of microorganisms takes place in phagocytic vacuoles called phagosomes [28]. The NADPH oxidase hereby is the major machinery that initially releases superoxide anions into the phagocytic space.

1.3 Proteolytic defence

1.3.1 Peptidases within PMN granules

Primary granules from neutrophils are the storage place for bactericidal peptides and proteins. Besides myeloperoxidase [29], lysozyme, defensins and permeability increasing proteins, the granules harbour the serine peptidases cathepsin G[6], neutrophil elastase[7] and proteinase 3[8] [28], [30]. Protein concentrations inside the granules of up to 500 mg/ml are reported [31]. Furthermore, secondary and tertiary granules contain iron and copper sequestering lactoferrin, lysozyme and gelatinase [32]. Interestingly, the antimicrobial α-defensin comprises 30-50% of the total protein in azurophilic granules

[6] Cathepsin G; CG
[7] Neutrophil elastase; NE
[8] Proteinase 3; PR3

[33]. Myeloperoxidase, which is considered a key constituent in the neutrophil cytotoxic armament, accounts for up to 5% of the total cell proteins [34].

Even though NE is not the most abundant peptidase, it is a key component regarding tissue remodelling, destruction and disease amplification. NE enhances tumor growth factor-β1[9] secretion via the MyD88 / IRAK pathway (see chapter 4) [35]. TGF-β1 contributes to the initiation of inflammatory events and plays important roles in tissue repair and fibrinolysis [36]. Since TGF-β1 is up-regulated in bronchial tissue from patients with severe asthma, NE is supposed to contribute to pathogenesis [37], [38].

1.3.2 Regulatory functions of neutrophil elastase and cathepsin G

Among all serine peptidases, NE is thought to be one of the major cytotoxic molecules because of its abundance and powerful degrading activity against a wide substrate variety. One of NE natural targets is elastin, an extremely stable component of the connective tissue. Elastin retains its integrity by boiling in 0.1 M NaOH at 95°C or successive one hour autoclave periods and can hardly be cleaved by any other peptidase except collagenase, cathepsin G and proteinase 3, although less effectively [39], [40]. Besides elastin, extracellular matrix proteins such as collagen, fibrin and proteoglycans, plasma proteins like coagulation factors and complement, IgG and IgM, peptidase inhibitors and various cytokines are natural targets of neutrophil elastase [41]. NE and CG even exert chemotactic functions by mediating the conversion of the biologically low activity precursor prochemerin to chemerin, a chemotactic protein for dendritic cells and macrophages [42]. NE-controlled proteolytic cleavage leads in most cases to loss of function, only complement and peptidases get activated.

Host tissues are protected from unregulated proteolytic activity by several soluble peptidase inhibitors. These include α_1-antipeptidase, secretory leukopeptidase inhibitor, elafin, α_2-macroglobulin and eglin [43], [44], [45], [46]. Nevertheless, NE uses an elaborate repertoire to circumvent this inhibition machinery: neutrophils in their tissue resident state are able to create a peptidase-inhibitor-free microenvironment that encompasses the surface to which they are adherent [47]. Secondly, α_1-antipeptidase and secretory leukopeptidase inhibitor are sensitive to oxidants released by neutrophils [48]. Thirdly, NE is relatively resistant to anti-peptidases if bound to elastin and similarly in a membrane bound form on the surface of activated neutrophils [49], [50]. The inactivation through oxidation is of particular significance in lung diseases of smokers [51]. Even though NE seems to be an inexorable peptidase, its destructive potential mainly becomes noticeable in pathological states (see section 1.4.3).

The difference between turnover rates of collagen and elastin remodelling at sites of tissue repair after injury or infection compared to normal turnover rates in healthy human beings can vary significantly. Elastase activity in wound healing processes and inflammatory responses is tightly controlled by the overwhelming amount of natural elastase inhibitors [43], [44], [45], [46]. However, several pathological states are known in which the uncontrolled cleavage of natural elastase target molecules causes severe collateral damage to various tissues.

[9] Tumor growth factor beta-1; TGF-β1

Human cathepsin G is found in the azurophilic granules of PMN. It is a serine peptidase with chymotrypsin activity [60]. It is an important enzyme modulator in inflammatory processes. Hereby, the low specific immune response elicited by PMN is linked to cells of the adapted immune system, in particular B cells, CD_4^+ and CD_8^+ T cells (CD = cluster of differentiation) and natural killer cells [61]. Furthermore, CG released from granulocytes binds to human natural killer cells and causes their enhanced cytotoxicity [62]. CG also acts as chemoattractant for circulating monocytes [63].

The imbalance between peptidases and their natural inhibitors may result in the uncontrolled digestion of most proteins of the extracellular matrix. Thus, CG, PR3, but foremost NE contribute to the development of connective tissue diseases like emphysema and rheumatoid arthritis [63], [64], [65]. However, α_1-antipeptidase, secretory leukopeptidase inhibitor, elafin and α_2-macroglobulin restrict and tightly control the digestive action of CG and other neutrophil serine peptidases [43], [44], [45], [46]. The numerous consequences of diseases that are linked to the dysregulation of CG and NE are briefly outlined in the following section.

1.3.3 Diseases

Interestingly, the cumulation of different pathological states mainly concerns the lungs. Several diseases such as asthma, cystic fibrosis, chronic obstructive pulmonary disease, emphysema, systemic inflammatory response syndrome, and more severe states thereof (acute lung injury, and adult respiratory distress syndrome), are directly or indirectly affected through the uncontrolled activity of NE, which is known to aggravate the state of the disease [52], [53], [54], [55], [56].

It is known that NE enhances TGF-β1 secretion via the MyD88 / IRAK pathway [35]. TGF-β1 contributes to the initiation of inflammatory events and plays important roles in tissue repair and fibrinolysis [36]. Furthermore, TGF-β1 release stimulates airway remodelling by inducing smooth muscle cell growth. The consequences of this process are fatal. The smooth muscle cells themselves are able to secrete various cytokines that might play a role in sub-mucosal inflammation and airway responsiveness, and they cause strengthened airway constriction leading to decreased cartilage volume. Such changes in the microvasculature are associated with increased airway vascularity, oedema, mucus production and the recruitment of inflammatory cells, which, of course, promote TGF-β1 release in a self-amplifying cycle [57]. Since TGF-β1 is up-regulated in bronchial tissue from patients with severe asthma, NE is supposed to contribute directly and indirectly to pathogenesis [37], [38].

Cystic fibrosis is a genetic disorder in which abnormal ion transport in the lungs leads to mucus accumulation predisposing patients to infection and neutrophilic inflammation [58]. The disease is further characterized by an increased level of NE and an imbalance between the metallopeptidases-8 and -9 and their natural inhibitor, named tissue inhibitor of metallopeptidase-1. This imbalance is based on proteolytic inactivation of the tissue inhibitor by NE and the activation of the precursor metallopeptidases [59].

NE, unlike other peptidases from granulocytes, is widely considered a key component regarding benign tissue remodelling on one side or malignant tissue destruction on the other side, and of course its involvement in a broad variety of

diseases. This circumstance raised great interest in medical and pharmacological studies, which tackled the search for powerful elastase inhibitors.

2. Rationale and aim of the project

The aim of this project is to find selective inhibitors against human neutrophil elastase and cathepsin G, starting from the scaffold of the naturally occurring sunflower trypsin inhibitor 1. The primary focus is to determine the structure-function relationship and the structural principles that govern specificity for this type of inhibitors.

REFERENCES

1.) Kuby, J., Immunology. fourth edition ed, ed. R.A. Goldsby, Kindt, T.J., Osborne, B.A. 2000, New York: Plenum Press.

2.) Roitt, I.M., Brostoff, J., Male, D.K., Kurzes Lehrbuch der Immunologie. third ed. 1995, Stuttgart, New York: Georg Thieme Verlag.

3.) Burg, N.D. and M.H. Pillinger, The neutrophil: function and regulation in innate and humoral immunity. Clin Immunol, 2001. 99(1): p. 7-17.

4.) Nathan, C., Neutrophils and immunity: challenges and opportunities. Nat Rev Immunol, 2006. 6(3): p. 173-82.

5.) Faurschou, M. and N. Borregaard, Neutrophil granules and secretory vesicles in inflammation. Microbes Infect, 2003. 5(14): p. 1317-27.

6.) Appelberg, R., Neutrophils and intracellular pathogens: beyond phagocytosis and killing. Trends Microbiol, 2007. 15(2): p. 87-92.

7.) Warren, M.K. and P. Ralph, Macrophage growth factor CSF-1 stimulates human monocyte production of interferon, tumor necrosis factor, and colony stimulating activity. J Immunol, 1986. 137(7): p. 2281-5.

8.) Chakraborty, A., et al., Granulocyte colony-stimulating factor promotes adhesion of neutrophils. Am J Physiol Cell Physiol, 2003. 284(1): p. C103-10.

9.) Takahashi, T., et al., Activation of human neutrophil by cytokine-activated endothelial cells. Circ Res, 2001. 88(4): p. 422-9.

10.) Batten, P., M.H. Yacoub, and M.L. Rose, Effect of human cytokines (IFN-gamma, TNF-alpha, IL-1 beta, IL-4) on porcine endothelial cells: induction of MHC and adhesion molecules and functional significance of these changes. Immunology, 1996. 87(1): p. 127-33.

11.) Polunovsky, V.A., et al., Induction of endothelial cell apoptosis by TNF alpha: modulation by inhibitors of protein synthesis. Exp Cell Res, 1994. 214(2): p. 584-94.

12.) Bloemen, K., et al., The allergic cascade: review of the most important molecules in the asthmatic lung. Immunol Lett, 2007. 113(1): p. 6-18.

13.) Adamko, D.J., et al., The rise of the phoenix: the expanding role of the eosinophil in health and disease. Allergy, 2005. 60(1): p. 13-22.

14.) Wong, C.K., et al., Intracellular signaling mechanisms regulating toll-like receptor-mediated activation of eosinophils. Am J Respir Cell Mol Biol, 2007. 37(1): p. 85-96.

15.) Marone, G., M. Triggiani, and A. de Paulis, Mast cells and basophils: friends as well as foes in bronchial asthma? Trends Immunol, 2005. 26(1): p. 25-31.

16.) Martin, C., et al., Chemokines acting via CXCR2 and CXCR4 control the release of neutrophils from the bone marrow and their return following senescence. Immunity, 2003. 19(4): p. 583-93.

17.) Suratt, B.T., et al., Neutrophil maturation and activation determine anatomic site of clearance from circulation. Am J Physiol Lung Cell Mol Physiol, 2001. 281(4): p. L913-21.

18.) Hengartner, M.O., The biochemistry of apoptosis. Nature, 2000. 407(6805): p. 770-6.

19.) Elbim, C., et al., Interleukin-18 primes the oxidative burst of neutrophils in response to formyl-peptides: role of cytochrome b558 translocation and N-formyl peptide receptor endocytosis. Clin Diagn Lab Immunol, 2005. 12(3): p. 436-46.

20.) Herrmann, J.M., et al., Sequential chemotactic and phagocytic activation of human polymorphonuclear neutrophils. Infect Immun, 2007. 75(8): p. 3989-98.

21.) Gewirtz, A.T. and E.R. Simons, Phospholipase D mediates Fc gamma receptor activation of neutrophils and provides specificity between high-valency immune complexes and fMLP signaling pathways. J Leukoc Biol, 1997. 61(4): p. 522-8.

22.) Grinstein, S., W. Furuya, and W.D. Biggar, Cytoplasmic pH regulation in normal and abnormal neutrophils. Role of superoxide generation and Na+/H+ exchange. J Biol Chem, 1986. 261(2): p. 512-4.

23.) Lazzari, K.G., P. Proto, and E.R. Simons, Neutrophil hyperpolarization in response to a chemotactic peptide. J Biol Chem, 1990. 265(19): p. 10959-67.

24.) Harakawa, N., et al., Random migration of polymorphonuclear leukocytes induced by GM-CSF involving a signal transduction pathway different from that of fMLP. J Leukoc Biol, 1997. 61(4): p. 500-6.

25.) Dinarello, C.A. and G. Fantuzzi, Interleukin-18 and host defense against infection. J Infect Dis, 2003. 187 Suppl 2: p. S370-84.

26.) Leung, B.P., et al., A role for IL-18 in neutrophil activation. J Immunol, 2001. 167(5): p. 2879-86.

27.) Smith, R.M. and J.T. Curnutte, Molecular basis of chronic granulomatous disease. Blood, 1991. 77(4): p. 673-86.

28.) Segal, A.W., How neutrophils kill microbes. Annu Rev Immunol, 2005. 23: p. 197-223.

29.) Corazza, A., et al., Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile Sulfolobus solfataricus. Proteins, 2006. 62(1): p. 64-79.

30.) Owen, C.A. and E.J. Campbell, The cell biology of leukocyte-mediated proteolysis. J Leukoc Biol, 1999. 65(2): p. 137-50.

31.) Reeves, E.P., et al., Killing activity of neutrophils is mediated through activation of proteases by K+ flux. Nature, 2002. 416(6878): p. 291-7.

32.) Bainton, D.F., Distinct granule populations in human neutrophils and lysosomal organelles identified by immuno-electron microscopy. J Immunol Methods, 1999. 232(1-2): p. 153-68.

33.) Lundy, F.T., et al., Antimicrobial activity of truncated alpha-defensin (human neutrophil peptide (HNP)-1) analogues without disulphide bridges. Mol Immunol, 2008. 45(1): p. 190-3.

34.) Klebanoff, S.J., Myeloperoxidase. Proc Assoc Am Physicians, 1999. 111(5): p. 383-9.

35.) Lee, K.Y., et al., Neutrophil-derived elastase induces TGF-beta1 secretion in human airway smooth muscle via NF-kappaB pathway. Am J Respir Cell Mol Biol, 2006. 35(4): p. 407-14.

36.) Sporn, M.B., et al., Transforming growth factor-beta: biological function and chemical structure. Science, 1986. 233(4763): p. 532-4.

37.) Ohno, I., et al., Transforming growth factor beta 1 (TGF beta 1) gene expression by eosinophils in asthmatic airway inflammation. Am J Respir Cell Mol Biol, 1996. 15(3): p. 404-9.

38.) Vignola, A.M., et al., Transforming growth factor-beta expression in mucosal biopsies in asthma and chronic bronchitis. Am J Respir Crit Care Med, 1997. 156(2 Pt 1): p. 591-9.

39.) Hay, E.D., Cell Biology of Extracellular Matrix. second edition ed. 1991, New York: Plenum Press.

40.) Lee, W.L. and G.P. Downey, Leukocyte elastase: physiological functions and role in acute lung injury. Am J Respir Crit Care Med, 2001. 164(5): p. 896-904.

41.) Kawabata, K., T. Hagio, and S. Matsuoka, The role of neutrophil elastase in acute lung injury. Eur J Pharmacol, 2002. 451(1): p. 1-10.

42.) Wittamer, V., et al., Neutrophil-mediated maturation of chemerin: a link between innate and adaptive immunity. J Immunol, 2005. 175(1): p. 487-93.

43.) Karnaukhova, E., Y. Ophir, and B. Golding, Recombinant human alpha-1 proteinase inhibitor: towards therapeutic use. Amino Acids, 2006. 30(4): p. 317-32.

44.) Moreau, T., et al., Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family. Biochimie, 2008. 90(2): p. 284-95.

45.) Wewers, M.D., D.J. Herzyk, and J.E. Gadek, Alveolar fluid neutrophil elastase activity in the adult respiratory distress syndrome is complexed to alpha-2-macroglobulin. J Clin Invest, 1988. 82(4): p. 1260-7.

46.) Feldman, S.R., S.L. Gonias, and S.V. Pizzo, Model of alpha 2-macroglobulin structure and function. Proc Natl Acad Sci U S A, 1985. 82(17): p. 5700-4.

47.) Rice, W.G. and S.J. Weiss, Regulation of proteolysis at the neutrophil-substrate interface by secretory leukoprotease inhibitor. Science, 1990. 249(4965): p. 178-81.

48.) Boudier, C. and J.G. Bieth, Oxidized mucus proteinase inhibitor: a fairly potent neutrophil elastase inhibitor. Biochem J, 1994. 303 (Pt 1): p. 61-8.

49.) Pham, C.T., Neutrophil serine proteases: specific regulators of inflammation. Nat Rev Immunol, 2006. 6(7): p. 541-50.

50.) Morrison, H.M., et al., Inhibition of human leukocyte elastase bound to elastin: relative ineffectiveness and two mechanisms of inhibitory activity. Am J Respir Cell Mol Biol, 1990. 2(3): p. 263-9.

51.) Cavarra, E., et al., Human SLPI inactivation after cigarette smoke exposure in a new in vivo model of pulmonary oxidative stress. Am J Physiol Lung Cell Mol Physiol, 2001. 281(2): p. L412-7.

52.) Brown, K.A., et al., Neutrophils in development of multiple organ failure in sepsis. Lancet, 2006. 368(9530): p. 157-69.

53.) Machen, T.E., Innate immune response in CF airway epithelia: hyperinflammatory? Am J Physiol Cell Physiol, 2006. 291(2): p. C218-30.

54.) Kuebler, W.M., Selectins revisited: the emerging role of platelets in inflammatory lung disease. J Clin Invest, 2006. 116(12): p. 3106-8.

55.) Barroso, B., N. Abello, and R. Bischoff, Study of human lung elastin degradation by different elastases using high-performance liquid chromatography/mass spectrometry. Anal Biochem, 2006. 358(2): p. 216-24.

56.) Fujishima, S., et al., Neutrophil elastase and systemic inflammatory response syndrome in the initiation and development of acute lung injury among critically ill patients. Biomed Pharmacother, 2007.

57.) Bergeron, C. and L.P. Boulet, Structural changes in airway diseases: characteristics, mechanisms, consequences, and pharmacologic modulation. Chest, 2006. 129(4): p. 1068-87.

58.) Rowe, S.M., S. Miller, and E.J. Sorscher, Cystic fibrosis. N Engl J Med, 2005. 352(19): p. 1992-2001.

59.) Gaggar, A., et al., Matrix metalloprotease-9 dysregulation in lower airway secretions of cystic fibrosis patients. Am J Physiol Lung Cell Mol Physiol, 2007. 293(1): p. L96-L104.

60.) Starkey, P.M., Barrett, A.J., Human cathepsin G. Catalytic and immunological properties. Biochem J., 1976. 155: p. 273.

61.) Hase-Yamazaki, T., Aoki, Y. , Stimulation of human lymphocytes by cathepsin G. Cell Immunol, 1995. 160: p. 24.

62.) Yamazaki, T., Aoki, Y.,, Cathepsin G enhances human natural killer cytotoxicity. Immunology, 1998. 93: p. 115-121.

63.) Miyata, J., Tani, K., Sato, K., Otsuka, S., Urata, T., Khagvaa, B., Furukawa, C., Sano, S., Sone, S., CatG: The significance in rheumatoid arthritis as a monocyte chemoattractant. Rheumatol. Int, 2007. 27(4): p. 375-382.

64.) Coakley, R.J., Taggart, C., O`Neill, S., McElvaney, N.G., Alpha-1-antitrypsin deficiency: biological answers to clinical questions. Am J Med Sciences, 2001. 321(1): p. 33-41.

65.) Abboud, R.T., Vimalanathan, S., , Pathogenesis of COPD. Part I. The role of protease - antiprotease imbalance in emphysema. Int J Tuberc Lung Dis, 2008. 12(4): p. 361-367.

Title: **Structural Basis of a potent Cathepsin G inhibitor derived from sunflower trypsin inhibitor 1**

Authors: Merz Tobias, Roschizki-Voser Heidi[1], Gompert Frank[2], Mittl Peer R.E.[1] und Grütter[*], Markus, G.

Affiliation: 1: Department of Biochemistry, University of Zürich, Winterthurer Str. 190, 8057 Zürich, Switzerland.

2: Polyphor Ltd., Allschwil, Switzerland.

*: to whom correspondence should be addressed.

Tel. +41-44-635 5580, Fax. +41-44-635 6834,

e-mail: gruetter@bioc.uzh.ch

Running title: Cathepsin G inhibitor derived from SFTI-1

Keywords: Cathepsin G, SFTI-1, elastase, inhibitor design, conformation

Abstract

The 14 amino acid long naturally occurring sunflower trypsin inhibitor (SFTI-1) scaffold is a popular, small and bi-cyclic peptide that has thoroughly been investigated with respect to its inhibitory potency and selectivity for trypsin, cathepsin G and human matriptase as well as its susceptibility for modifications. In this work, SFTI-1 was redesigned into a potent cathepsin G inhibitor (K_i = 11 nM, POL4634) and a neutrophil elastase inhibitor (K_i = 35 nM, POL4957). They were crystallized in complex with cathepsin G, porcine pancreatic elastase and human neutrophil elastase. The structural analysis in combination with inhibitory data suggests that imposed conformational restraints of the peptidase have more impact on the inhibitory activity of the non-covalently binding peptidomimetic inhibitors than measurable inhibitor-peptidase-interaction parameters.

Introduction

Naturally occurring peptidase inhibitors are conformationally restricted and show remarkable stability towards heat, proteases and denaturing agents [Makhatadze, 1993], [Colgrave, 2004]. These beneficial properties can often be attributed to disulfide cross-linking or backbone cyclization, which compensate the lack of a hydrophobic core. The pancreatic trypsin inhibitors from the Kunitz- and Kazal-types, the soybean Bowman-Birk inhibitors and the cyclotides [Makhatadze, 1993], [Qi, 2005], [Avrutina, 2005], [Craik, 1999] are molecules that show these properties. Disulfide bridges and hydrogen bonds rigidify the fold and restrict the canonical peptidase-binding loop into an extended conformation forming an anti-parallel β-sheet in the proper conformation to facilitate hydrogen bond formation within the enzyme binding site [Laskowski, 2000], [Apostoluk, 1998], [Tyndall, 2005].

The bi-cyclic trypsin inhibitor from sunflower seeds (SFTI-1) consists of 14 amino acids and folds into a short anti-parallel β-hairpin structure. SFTI-1 is cross-linked by a single disulfide bridge and its binding loop shows high sequence and structural similarity to Bowman-Birk inhibitors [Luckett, 1999]. The stability and small size of SFTI-1 as well as the availability of structural data makes it an interesting molecular template for drug design applications. SFTI-1 exhibits a strong inhibitory activity against trypsin K_i = 0.3 – 1x10^{-10} M [Luckett, 1999], [Daly, 2006], [Garcia Boy, 2009] and cathepsin G (K_i < 1.5x10^{-10} M) [Luckett, 1999] but only very weak affinity against elastase (K_i ~ 1x10^{-4} M) [Luckett, 1999]. Several studies show, that SFTI-1 is well suited as a template for the design of β-hairpin peptidomimetics, maintaining strong inhibitory activity against trypsin, cathepsin G or matriptase [Descours, 2002],[Legowska, 2009], [Li, 2007]. The serine peptidases cathepsin G, human leucocyte elastase and proteinase 3 are secreted by neutrophils at sites of

inflammation. They play a critical role in the activation of cytokines and growth factors and stimulate various cellular receptors (reviewed in [Wiedow, 2005]).

The focus of our study was the structural analysis of a redesigned SFTI-1 inhibitor in complex with cathepsin G, human neutrophil elastase and porcine pancreatic elastase. Kinetic experiments showed strong inhibitory potency against cathepsin G and neutrophil elastase while trypsin could not be inhibited with the modified inhibitors. The redesign of SFTI-1 included four changes: (i) Ile7 was changed to Asp7, because cathepsin G, in contrast to trypsin, revealed a patch of positive charge density around the S2` binding site. (ii) The β-turn forming tripeptide Pro13-Asp14-Gly1 in the non-prime site of SFTI-1 was replaced by the shorter ^{D}Pro13-^{L}Pro1 dipeptide with the aim to improve the stability and to lock the inhibitor in a favorable conformation. (iii) Arg2 in P4 was replaced by Phe2 to satisfy the requirements of the hydrophobic S4 pocket of cathepsin G. (iv) Lys5 in P1, the selectivity discriminating residue for trypsin-like peptidases [Daly, 2006], [Descours, 2002], was substituted with the non-standard aliphatic amino acid norleucine (Figure 1). Norleucine lacks the terminal ε-amino group and disables the formation of a selectivity determining salt bridge at the bottom of the S1 pocket. The cyclic and non-covalent cathepsin G inhibitor POL4634 finally consisted of 13 amino acids including the two non-natural amino acids norleucine and D-proline at positions 5 and 13, respectively.

We determined the crystal structures of the SFTI-1 derivative POL4634 in complex with cathepsin G and porcine pancreatic elastase and of the SFTI-1 derivative POL4957 in complex with human neutrophil elastase and analyzed the respective binding interfaces. These complex structures not only revealed the necessary structural features of the inhibitors that are required for binding, but also the necessary structural predisposition of the binding site. The conformation of the inhibitor hereby plays an essential role leading to decreased affinity when

the binding site of the enzyme imposes a conformational strain. Moreover, we were able to solve the structure of POL4634 bound to PPE in three different states (i) in its cyclic form, (ii) in its cyclic form before nucleophilic attack and (iii) in a cleaved state that represents a stable acyl-enzyme intermediate state.

Results

1. Crystal structure of POL4634 in complex with cathepsin G.

The crystal structure of cathepsin G in complex with POL4634 was determined at 1.8 Å. The superposition of this complex structure and cathepsin G in complex with Suc-Val-Pro-Lys^{p}-$(Oph)_2$ (PDB-ID: 1AU8) resulted in an rmsd of 0.35 Å for all Cα-atoms except for residues 36 to 38 and indicated that both structures were almost identical. The crystallographic data are shown in Table 1. The inhibitor POL4634 formed a distorted anti-parallel β-sheet that was stabilized by six intra-molecular hydrogen bonds, involving main-chain atoms from residues Phe2I, Thr4I, Pro8I, Ile10I, and Tyr12I as well as side chain atoms from Thr4I and Ser6I. In addition, a disulfide bridge between Cys3I and Cys11I cross-linked the adjacent β-strands (Figure 2A). However, the increased distance between Cys3I-Sγ and Cys11I-Sγ (3.1 Å) and the alternative side chain conformation of Cys3I suggested a partial opening of this disulfide bridge due to radiation damage. The cyclic POL4634 β-hairpin contained two loops that were located in the prime- and non-prime cathepsin G binding pockets. The prime-site loop was stabilized by the ^{L}Pro8-^{L}Pro9 dipeptide and adopted a cis-peptide conformation comparable to the naturally occurring SFTI-1 inhibitor in complex with trypsin [Luckett, 1999]. The binding interface between POL4634 and cathepsin G included eleven hydrogen bonds, a surface area of 668.6 Å^2 and a shape complementarity (SC) of 0.677 (Table 2). The POL4634 inhibitor occupied the cathepsin G S4 to S2' binding pockets (Figure 2A). The intra-molecular β-sheet conformation of POL4634 predefined the orientation of the residues Pro1I to Nle5I, which interacted with residues Ser218 to Ser214 from cathepsin G in the canonical anti-parallel β-sheet conformation [Garavilla de, 2005]. Phe2I (P4) T-stacked on the side chain of Phe172 and together with Tyr12I, they were involved in van der Waals interactions with hydrophobic residues in the non-prime-site pocket. Cys3I (P3), Thr4I (P2) and Ser6I (P1`) formed main

chain and side chain hydrogen bonds and stabilized the central part of POL4634 while Nle5I occupied the S1 binding site. Asp7I (P2`) was surrounded by the side chains of Arg41, Arg143, and Lys192 of the S2` binding pocket (Figure 2A). However, Asp7I did not directly interact with any of these positively charged side chains but formed a water-mediated salt bridge with Arg143. The residues Pro8 to Pro13 faced the solvent. The electron density of norleucine suggested a planar, sp^2 hybridized Nle5I-C atom. Its carbonyl oxygen made hydrogen bonds with the backbone amide groups of Gly193, Asp194 and Ser195 that form the oxyanion hole.

Ser40 of cathepsin G was located in a loop region that adopted substantially different conformations in the complexes with POL4634 and Suc-Val-Pro-Lys^p-$(Oph)_2$ (PDB-ID: 1AU8). Whether these different conformations could be attributed to inhibitor binding remained unclear, because residues 36B to 38 were involved in a crystal contact. Pro36B-O formed a hydrogen bond with Gly173-N (2.8 Å) from the symmetry related molecule (generated by the symmetry operator –X+1/2, -Y, Z+1/2). The superposition of the complex structures of the macrocyclic inhibitor POL4634, the tripeptidic inhibitor Suc-Val-Pro-Lys^p-$(Oph)_2$ and the aliphatic inhibitor JNJ-10311795 [Zani, 2009] (PDB-ID: 1T32) in complex with cathepsin G revealed conformational differences in the S1 pocket. The replacement of lysine with a norleucine in P1 caused a shift of the Glu226 side chain at the bottom of the S1 pocket, because Glu226 could no longer form a polar interaction with the side chain of the inhibitor. In addition, the active site of cathepsin G in complex with hydrophobic ligands such as norleucine or the naphthalene moiety of the JNJ-10311795 inhibitor (Figure 3) did not harbor water molecules.

2. Crystal structure of POL4634 in complex with porcine pancreatic elastase.

The crystal structure of PPE in complex with POL4634 was determined at 1.8 Å resolution with three PPE/POL4634 complexes in the asymmetric unit. The crystallographic data are shown in Table 1. Similarly to the complex structure of cathepsin G, POL4634 in complex with PPE formed a distorted anti-parallel β-sheet that was stabilized by intra- and intermolecular hydrogen bonds and a disulfide bond between Cys3I and Cys11I. Details of the inhibitor-enzyme-interaction parameters are listed in Table 2. The three PPE/Pol4643 complexes revealed substantial differences at the P1 site (Figure 4). The electron density contoured at 1.0σ around P1 showed the inhibitor bound in three different states: (i) an intact peptide bond of chain L (Figure 4B), (ii) a partially sp3 hybridized peptide bond of chain M (Figure 4C) and (iii) a broken peptide bond of chain N featuring an acyl enzyme intermediate state between the active site SerC195 and the backbone of NleN5 (Figure 4D). (A, B and C refer to the peptidase chains; L, M and N refer to the corresponding inhibitor chains). A comparison of the inhibitor-enzyme-interaction parameters reflected the different states of the inhibitor chains: Whereas the closed chains L and M revealed more intermolecular hydrogen bonds and a significantly higher shape complementarity than the opened chain N, they displayed a lower number of intramolecular hydrogen bonds and a smaller buried interaction surface (Table 2). POL4634 acts as a reversible non-covalently binding inhibitor for PPE and cathepsin G and represents an enzyme-substrate inhibitory complex, as it was described by Luckett et al. for SFTI-1 in complex with trypsin [Luckett, 1999]. In contrast to cathepsin G, where POL4634 features an enzyme-substrate inhibitory complex without connecting electron density between norleucine in P1 and the active site Ser195, the three NCS related PPE/POL4634 copies reveal three different states of the inhibitor cleavage process.

3. Crystal structure of POL4957 in complex with human neutrophil elastase.

The crystal structure of HNE in complex with POL4957 was determined at 1.75 Å resolution and contained one *apo*-enzyme and two *holo*-enzymes in the asymmetric unit. Superpositions of the structures with and without POL4957 in comparison to HNE (PDB-ID 1HNE) showed rmsd values between 0.200 and 0.435 indicating that the structures were almost identical. However, the residues Leu143 to Ile151 formed a loop, which assumed slightly different conformations in the three molecules of the asymmetric unit. This loop of a symmetry related molecule (symm. op. X + 0.5, Y + 0.5, Z + (1 0 0)) is involved in crystal contacts with the active site pocket of the *apo*-enzyme and prevented binding of the inhibitor. The inhibitor POL4957 formed a distorted anti-parallel β-sheet that was stabilized by seven intra-molecular hydrogen bonds, involving main-chain atoms from residues Nle2I, Thr4I, Pro8I, Gln10I, and Tyr12I as well as side chain atoms from Thr4I and Ser6I. Cys3I and Cys11I cross-linked the adjacent β-strands with a disulfide bond (Figure 2B). The binding interface of POL4957 and the peptidase included a surface area of 613.7 $Å^2$ and 620.1 $Å^2$, a shape complementarity (SC) of 0.702 and ten and eight hydrogen bonds for the inhibitor chains I and J, respectively (Table 2). The inhibitor POL4957 occupied the HNE S4 to S2` binding pockets and interacted with the residues Ser214 to Gly218 of HNE in the canonical anti-parallel β-sheet conformation. Nle2I (P4) is in van der Waals contact with Leu99B and Phe215 in the non-prime-site pocket. Cys3I (P3), Thr4I (P2) and Ser6I (P1`) formed main chain and side chain hydrogen bonds and stabilized the central part of POL4957 while Ala5I (P1) pointed into the S1 binding site. The long aliphatic side chain of oGly7I (octyl-glycine) (P2`) faces the solvent. However, its electron density was too weak to clearly identify interacting residues of the hydrophobic prime-site groove. The amide backbone of Ala5I was involved in four hydrogen bonds with Ser214 and the amide groups of Gly193, Asp194 and Ser195 that form the oxyanion hole. Although the distances of Ser195-Oγ to the Ala5I carbonyl atom were close (2.46 Å and 2.54

Å of chains I and J, respectively), the electron density of alanine in P1 suggested a planar, sp^2 hybridized carbonyl atom.

4. Kinetic analysis of the inhibition of POL4634 and POL4957 on cathepsin G, elastases and trypsin

The inhibitory potency of POL4634 and POL4957 was tested against human cathepsin G, porcine pancreatic elastase (PPE), human neutrophil elastase (HNE) and bovine trypsin (Figures 5A-D and 6A-D). The substrates used for monitoring the activity of cathepsin G, PPE and trypsin were Suc-Ala-Ala-Pro-Phe↓AMC, Suc-(OMe)-Ala-Ala-Pro-Val↓AMC and Boc-Gln-Ala-Arg↓AMC respectively. Arrows in the substrate acronyms indicate the scissile bonds. The amino-methyl-coumarin (AMC) leaving group was detected fluorometrically at $\lambda_{ex\ 360}/\lambda_{em\ 465}$. Initial inhibition tests with POL4634 and POL4957 showed that the formation of product in the presence of inhibitor was linear from the beginning of the reaction. Neither POL4634 nor POL4957 inhibited trypsin at a concentration of 16 uM, corresponding to a 9000 fold inhibitor excess (Figure 5D and 6D). For cathepsin G, HNE and PPE, initial velocities at variable substrate and inhibitor concentrations were treated as steady-state rates. The best fit model for cathepsin G inhibition by POL4634 with Suc-Ala-Ala-Pro-Phe↓AMC at pH 6.5 was linear, competitive tight binding with a K_i = 11nM (Figure 5A). A K_i = 35 nM was determined for HNE using POL4957 (Figure 6B). The cross-inhibition of HNE and PPE with POL4634 and of cathepsin G with POL4957 was weak (Figures 5B, 5C and 6A). The specific velocity plots, typical for linear competitive inhibition, are shown in the inset of Figures 5 and 6. PPE inhibited with POL4634, initial velocity data were fitted to equation 3 (materials and methods). Figure 2C presents POL4634 in complex with cathepsin G and PPE and SFTI-1 in complex with trypsin in a superposition. Table 3 summarizes the inhibitory

data. The kinetic analysis classifies POL4634 as a compound with excellent inhibitory potency for cathepsin G.

5. Cross comparison of inhibitors

POL4634 inhibited cathepsin G (K_i = 11 nM) and PPE (K_i = 2.2 uM) but not HNE. POL4957, inhibited PPE (K_i = 21.4 nM) and HNE (K_i = 35 nM), cathepsin G, however, was only weakly inhibited (K_i > 6uM). In order to explain the differences in inhibitory potency between POL4634, POL4957 and the peptidases, we modeled POL4634 on the basis of POL4957 and assumed that this model inhibitor is able to assume the conformation of POL4957 bound in the active site pocket of HNE. Hence, Nle2I was replaced with phenylalanine, Ala5I with norleucine, oGly7I with aspartate and Gln11I with isoleucine. Correspondingly, POL4957 was modeled on the basis of POL4634 into the active-site pockets of cathepsin G and PPE.

Modeled POL4634 in the active site of HNE suggested sterical clashes of phenylalanine (2I) with Arg191 in the non-prime-site pocket and of norleucine (5I) in the active-site pocket. Aspartate (7I) and isoleucine (11I) were not in van der Waals contact with residues of the active-site pocket. Modeled POL4957 in the active sites of cathepsin G and PPE did not indicate any sterical clashes between the exchanged inhibitor residues and the active site pockets. Norleucine (2I) and alanine (5I) accommodated well within the hydrophobic non-prime-site pockets and the active sites. While aspartate could make water-mediated salt bridges with Arg41 and Arg143 in the prime-site pocket of cathepsin G, the aliphatic octyl chain was expected to be highly unfavorable in this patch of positive charge density. However, octyl-glycine could well accommodate within the hydrophobic prime-site groove of PPE. The replacement of isoleucine (11I) by glutamine suggested additional hydrogen bonds with Gln96 in cathepsin G and van der Waals contacts in PPE.

These structural findings are well in agreement with our inhibition studies: Neutrophil elastase is not inhibited by POL4634 due to expected sterical clashes of norleucine in the active site pocket and phenylalanine (P4) in the non-prime-site pocket. Cathepsin G is only weakly inhibited by POL4957 because of highly unfavorable interactions of the octyl-side chain with the positively charged S2` binding site. PPE is inhibited by POL4634 and POL4957. Both inhibitors do not clash within the active site of PPE. The loss of hydrophobic interactions in the S1 binding site due to the exchange of norleucine by alanine may be supercompensated by the gain of hydrophobic interactions through the aspartate-octyl-glycine exchange in P2` resulting in 100 fold lower K_i of POL4957.

6. Binding modes of POL4634 and POL4957

The superposition of cathepsin G and PPE revealed the structural features that accounted for the different shapes of the active sites. (i) The non-prime-site loop of PPE contained a serine residue (Ser217), which displaced arginine (Arg217A) from where the corresponding lysine residue in cathepsin G (Lys217) was located; (ii) the prime-site loop, comprising residues 32 to 42, was perpendicularly oriented relative to the active site cleft of cathepsin G, giving the groove a curved, narrow and deep appearance. In PPE however, this loop pointed away from the active site cleft and was almost parallel oriented relative to the active site cleft, which gave the groove a more elongated, wider and shallower appearance (Figure 7A). Consequently, POL4634 bound in an extended conformation in PPE and in a curved conformation (Figure 7B) in the active site cleft of cathepsin G without significant differences in the interface interaction parameters.

Discussion

The high resolution structure of POL4634 in complex with cathepsin G and kinetic inhibitory data illustrate the successful redesign from a potent trypsin to a potent cathepsin G inhibitor. This was mainly achieved by the substitution of lysine to norleucine in P1, which prevents the formation of a salt bridge in the active site pocket and the substitution of arginine to phenylalanine in P4, which better accounts for the hydrophobic S4 site. These major mutations completely abolish trypsin inhibition and result in a potent cathepsin G inhibitor (K_i = 11 nM). The replacement of the β-turn forming Pro13-Asp14-Gly1 with the ^{D}Pro-^{L}Pro motif does not affect the trypsin inhibition. This is a surprising finding, since Descours et al. showed that the ^{D}Pro-^{L}Pro SFTI-1 variant displayed even stronger inhibitory potency towards trypsin [Descours, 2002].

The inhibitory potency of POL4634 against cathepsin G and PPE and POL4957 against HNE agrees very well with previously reported inhibition studies of SFTI-1 derivatives. The P1 lysine has been shown to be highly important as a selectivity determining feature for trypsin inhibition. Its exchange to alanine in SFTI-1 variants significantly decreased the affinity from $3x10^{-11}$ M to $1.9x10^{-7}$ M [Daly, 2006], or in shorter inhibitor variants from $1.03x10^{-7}$ M to $>1x10^{-5}$ M [Descours, 2002]. The exchange of lysine to phenylalanine of another SFTI-1 variant was shown to decrease the affinity towards cathepsin G [Legowska, 2010]. In addition to lysine in P1, arginine in P4 of SFTI-1 explores a hydrophobic area of the active site clefts of PPE, cathepsin G and trypsin. Hence, the replacement of the P4 arginine by valine increased the inhibitory potency towards PPE from $3.5x10^{-5}$ M to $3x10^{-7}$ M [Hilpert, 2005], and the replacement by alanine decreased the inhibitory potency towards trypsin from $3x10^{-11}$ M to $4x10^{-10}$ M [Daly, 2006]. Taking these data into account, we concluded that the combination of norleucine in P1 and phenylalanine in

P4 of the SFTI-1 variant POL4634 is mainly responsible to ensure potent cathepsin G inhibition and to abolish trypsin inhibition.

The kinetic data showed that POL4634 is an inhibitor for PPE in the high nanomolar range. However, the structural comparison of the binding interfaces of PPE and cathepsin G does not explain the difference of almost two orders of magnitude in affinity. Intra-molecular hydrogen bond patterns, sizes of the buried surface areas and shape complementarities are similar (Table 2) and the overall curved shapes of the active sites enables POL4634 to form an extended inter-molecular hydrogen bonding pattern in the canonical anti-parallel β-sheet conformation in both peptidases. To fit into the active site of PPE, however, POL4634 must adopt a more straight conformation. The prime site and more pronounced the non-prime site flanking loops (residues 32 to 42 and 216 to 218, respectively) narrow and apparently straighten the groove which makes it impossible for POL4634 to bind in a similarly curved manner as in cathepsin G (Figure 7A). Therefore, we propose that the difference in affinity must originate from a conformational restraint imposed by the peptidase. The bent conformation of POL4634 bound to cathepsin G as well as the similarly bound POL4957 inhibitor in the active site of HNE supposedly represent the energetically most favorable conformation of these inhibitors while an elongated conformation in the case of PPE suggests an energetically less favorable conformation, loss in entropy and thereby a decreased inhibitory potency. This hypothesis is supported by the crystal structure of SFTI-1 in complex with trypsin [Luckett, 1999] and the NMR structure of SFTI-1 (PDB ID: 1JBN). The solution structure and the complex structure show the same inhibitor conformation, which suggests that SFTI-1 is bound to trypsin in its energetically most favorable conformation. Superimposing SFTI-1 onto POL4634 reveals a similarly bent conformation of POL4634 in

complex with cathepsin G and therefore suggests a favorable conformation (along with high inhibitory potency) in the complex structure.

A number of naturally occurring or engineered elastase (PPE and HNE) specific inhibitors are known to date (SLPI, elafin, MCoTI-II or greglin from *Schistocerca gregaria*) [Zani, 2009], [Thongyoo, 2009], [Brillard-Bourdet, 2006]. The structures of OMTKY3 and 1/2SLPI in complex with HNE (PDB-ID: 1PPF and 2Z7F, respectively), elafin in complex with PPE (PDB-ID: 1FLE) all show a scaffold protein with a disulfide bridge stabilized large binding loop. Apparently, no conformationally constrained β-turns flank the inhibitor peptide stretch that interacts with the enzymes. This feature allows for flexibility and adaptability while little conformational strain is imposed by the enzymes` active site on the inhibitor upon binding.

POL4634 in the cleaved form in contrast to the non-cleaved form suggests a poorer fit in the active site groove of PPE. Although the interaction area is larger, it displays only a few intermolecular hydrogen bonds, and a lower shape complementarity value compared to the non-cleaved form. Thus, the cleavage event by PPE shifts the enzyme-inhibitor complex from a high-energy “closed” state, representing a tight binding mode, to a low-energy “open” state (cleaved peptide bond, representing a loose binding mode. We speculate that this conformational relaxation of POL4634 promotes the dissociation from the peptidase. Although no signs of inhibitor cleavage were observed, we cannot exclude that POL4634 is not cleaved by cathepsin G over a long period of time.

Summary

In summary, we conclude that the inhibitor potency does not exclusively depend on extensive inter-molecular hydrogen bonding networks, interaction areas and shape complementarities but, as a superior criterion, on a conformational strain on the inhibitor imposed by the peptidase. The structural study of the redesigned SFTI-1-based inhibitor POL4634 in complex with cathepsin G and PPE in combination with inhibitory data clearly illustrates that the lowest energy conformation of the free inhibitor is an important determinant for inhibitor potency. The future redesign of SFTI-1-based inhibitors must consider the conformational states by providing more flexibility to the frameworks which better allow adapting to the given spatial arrangement of the active sites of elastases and other enzymes.

TABLES

Table 1: Data and refinement statistics

Data	***PPE***	***Cathepsin G***
space group	P321 (145)	P212121 (19)
cell parameters	a = b = 95.1; c = 70.1 α = β = 90; γ = 120	a = 39.8; b = 127.0; c = 43.5 α = β = γ = 90°
resolution	19.75 – 1.8	43.5 – 1.8
obs. Reflections	65110	20946
completeness (%)	99.1 (98.8)[a]	98.4 (83.9)[a]
redundancy	5.64	6.68
Rsym (% on I)	8.3 (43.5)[a]	10.6 (34.3)[a]
I/σI	14.89 (4.55)[a]	13.03 (4.64)[a]
Refinement		
molecules / AU	3	1
R / Rfree (%)	15.5 / 19.5	17.8 / 21.8
waters	645	201
bond length (Å)	0.017	0.010
bond angles (°)	1.827	1.242
residues in most favoured regions	651 (88.0%)	220 (97.8%)
generally allowed regions	82 (11.1%)	5 (2.2%)
disallowed regions	7 (0.9%)	0
twinning fraction	0.47	
twin law	h,-h-k,-l	

[a]: refers to resolution shell 1.8 – 1.9 Å

Data for HNE to be appended

Table 2: Summary: interactions of inhibitors and peptidases

	HB Intermol.	***HB Intramol.***	***SC***	***BSA (Å^2)***
CatG	11	6	0.677	668.6
PPE (A/L)	9	3	0.748	735.6
PPE (B/M)	13	5	0.752	728.9
PPE (C/N)	6	9	0.681	793.1
HNE (I)	10	7	0.702	613.7
HNE (J)	8	7	0.702	620.1
Trypsin (SFTI-1)	11	8	0.731	690.1

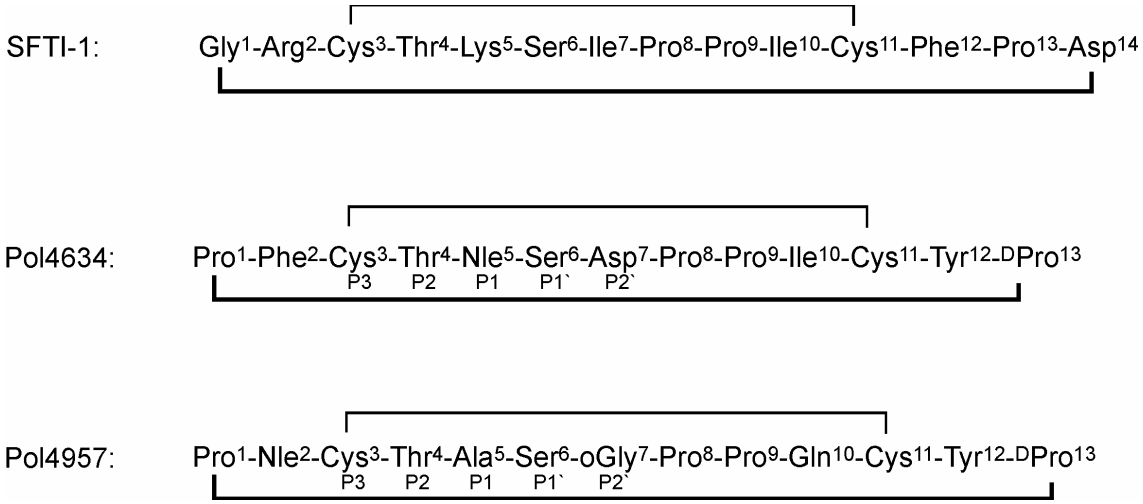
SC: shape complementarity; BSA: buried surface area

Table 3: Summary of inhibitory data for cathepsin G, neutrophil elastase, pancreatic elastase and trypsin

	POL4634	***POL4957***
Cathepsin G	K_i = 11 nM ±0.6	K_i > 6000 nM
Neutrophil elastase	Weak inhibition	Ki = 35 nM ±5
Pancreatic elastase	K_i = 2190 nM ±67	K_i = 21 nM ±1.5
Trypsin	No inhibition detected	No inhibition detected

Figures and Figure captions

Figure 1)



Primary sequences with superscript numbering of the sunflower trypsin inhibitor SFTI-1, POL4634 and POL4957. Thin lines denote cysteine bridges, strong lines denote a covalent link between the first and last amino acid. Positions around the P1 active site pocket residue are indicated.

Figure 2A

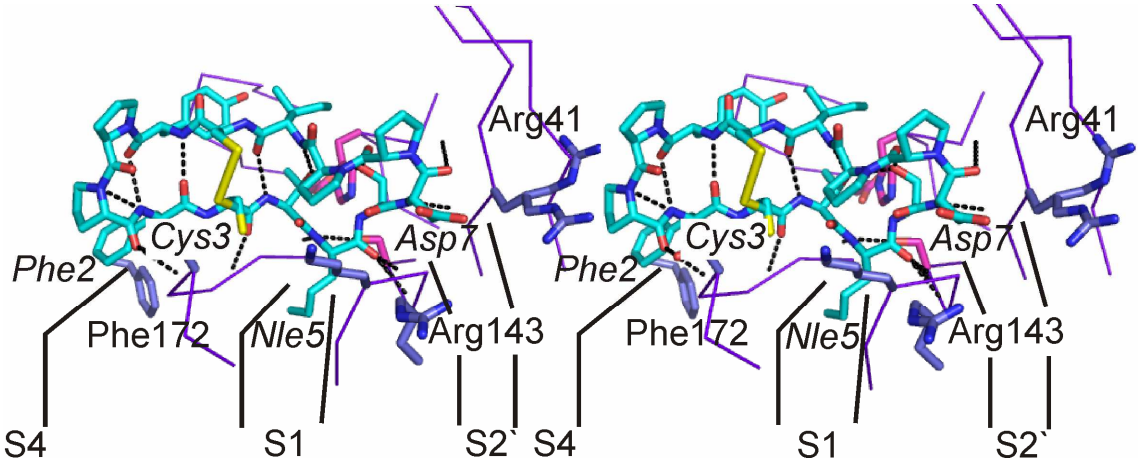


Figure 2B)

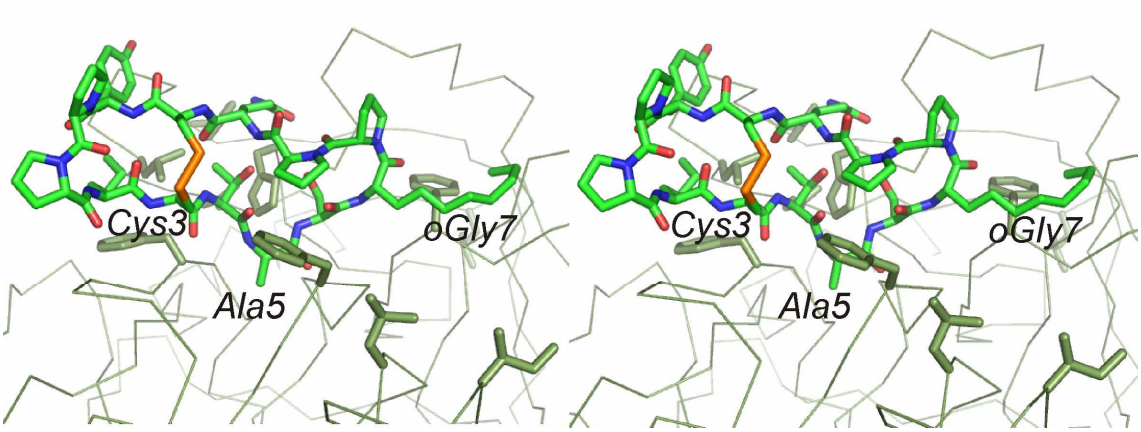


Figure 2C)

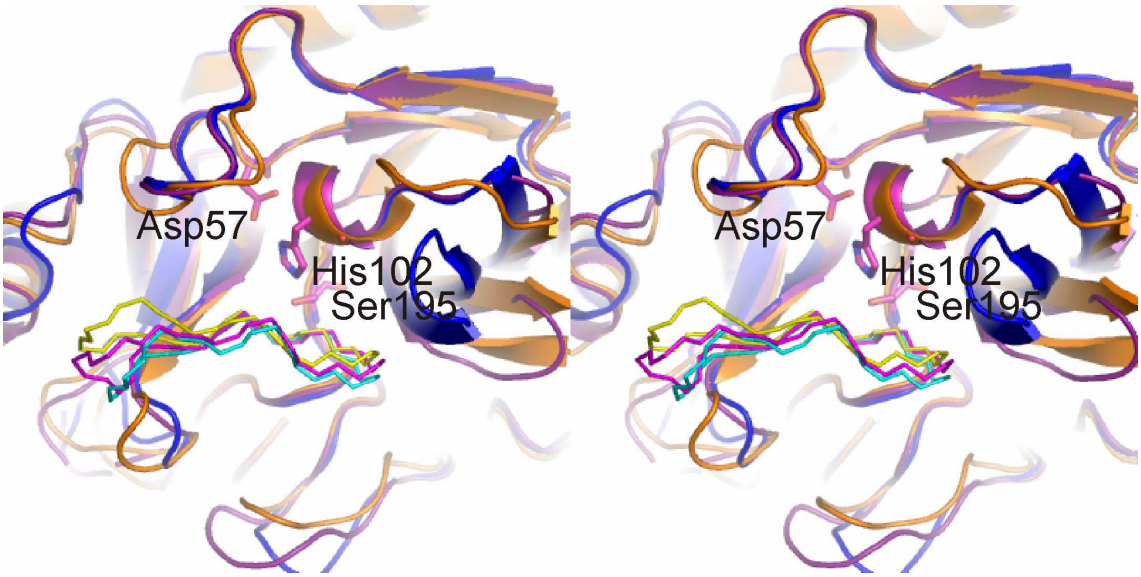


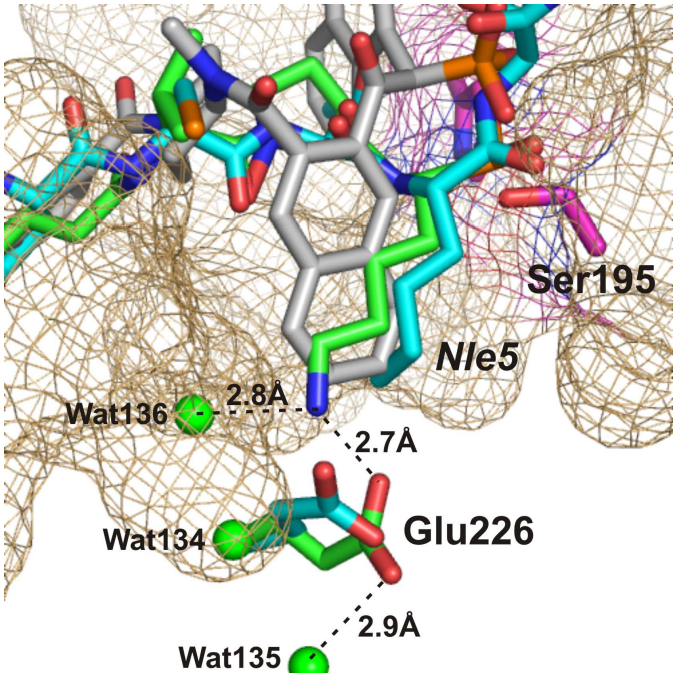
Figure 2)

A) POL4634 in light blue shown as sticks in complex with cathepsin G (dark blue) shown as ribbon representation. Hydrogen bonds are indicated with dashed lines. Phe2, Nle5 and Asp7 point into the S4, S1 and S2` pockets, respectively, indicated with solid lines. Residues of POL4634 are labelled in italics.

B) POL4957 in light green shown as sticks in complex with human neutrophil elastase (dark green) shown as ribbon representation. Hydrophobic residues of HNE within the active site groove are shown as sticks.

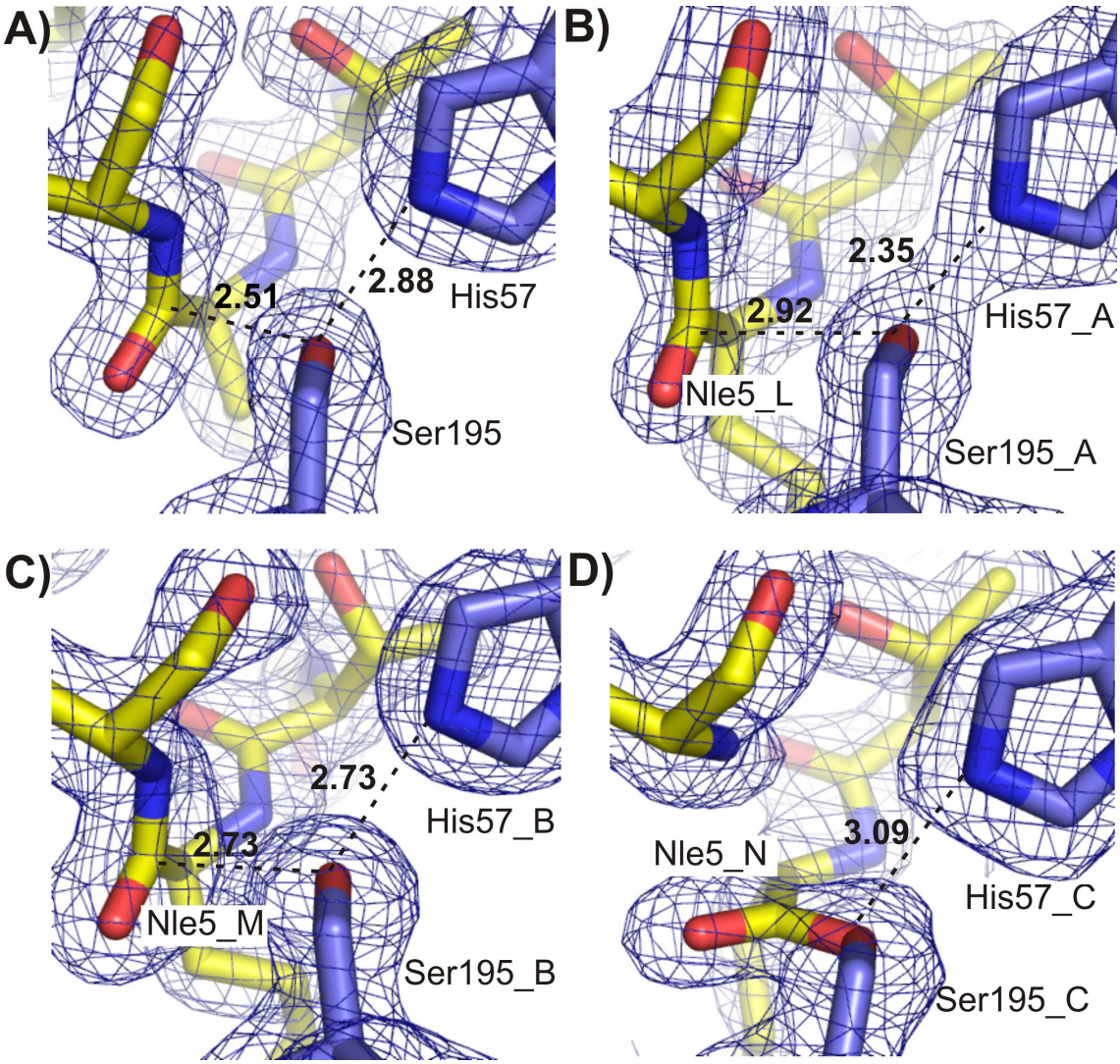
C) Superposition of POL4634 in complex with cathepsin G and PPE and SFTI-1 in complex with trypsin. Cathepsin G, PPE and trypsin are shown in cartoon representation in dark blue, orange and magenta, respectively. The inhibitor chains are represented as lines with their backbone atoms coloured in light blue, yellow and magenta, respectively. The catalytic triad of trypsin (Asp57, His102 and Ser195) is shown as sticks.

Figure 3)



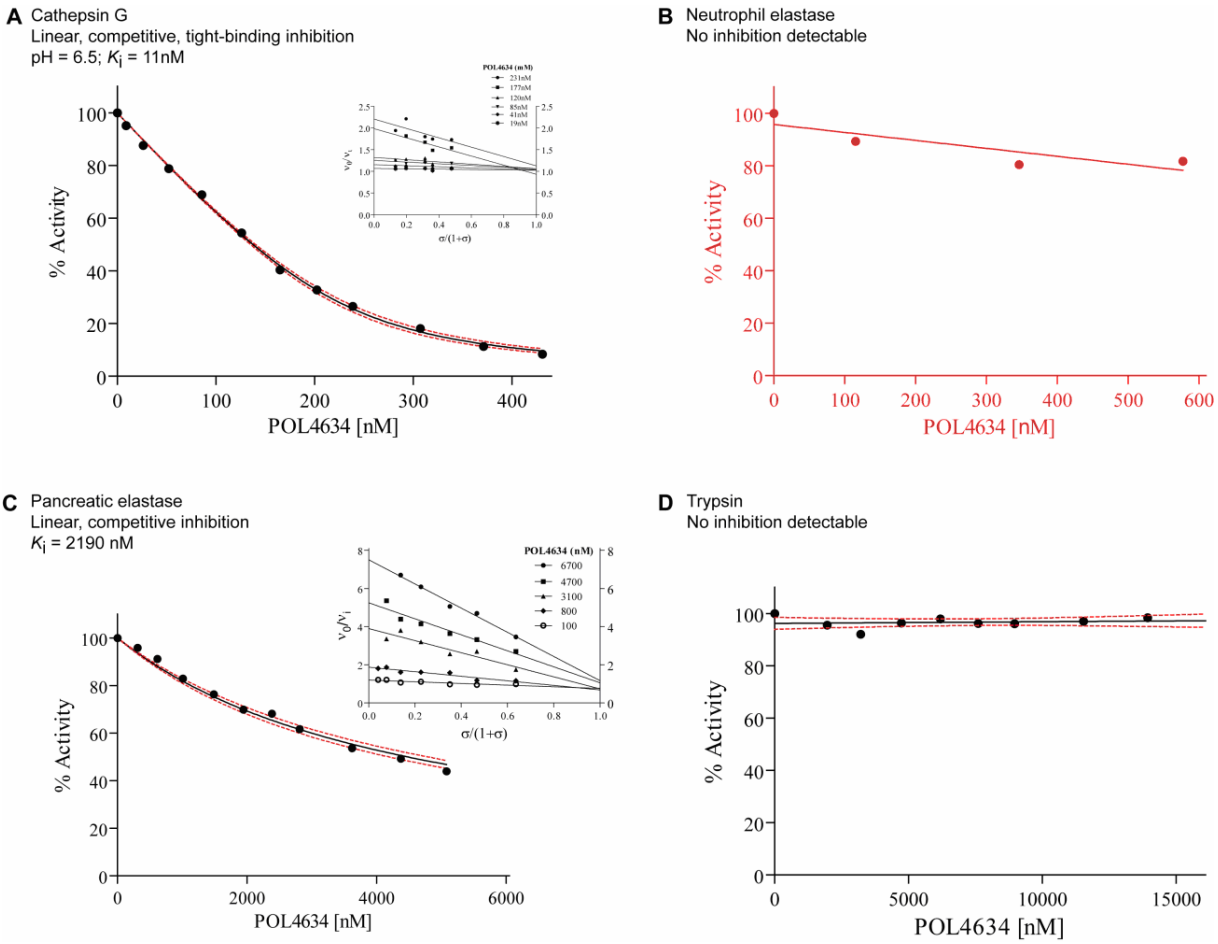
S1 pocket of cathepsin G with superimposed inhibitors. POL4634 is depicted in light blue, Suc-Val-Pro-Lys^{P}-$(OPH)_2$ in green (PDB-ID: 1AU8) and JNJ-10311795 in grey (PDB-ID: 1T32). The green water molecules Wat134 – Wat136 belong to PDB-ID 1AU8 and make hydrogen bonds to the ε-amino group of the phosphonate-lysine of Suc-Val-Pro-Lys^{P}-$(OPH)_2$-inhibitor and Glu226. Glu226 can adopt different conformations in combination with a salt bridge forming (lysine) or hydrophobic (norleucine) P1 residue.

Figure 4)



Close-up view into the scissile Nle5-Ser6 peptide bond of POL4634 bound to cathepsin G and PPE. A, B) The peptide bonds of POL4634 in complex with cathepsin G (A) and PPE (B) are intact, no sign of sp3-hybridization. C) Connecting density between NleM5-C and Ser195-Oγ of POL4634 in complex with second PPE molecule suggests partial sp3-hybridization. D) Acyl-enzyme intermediate between open POL4634 and the active site serine. Hydrogen bonds are indicated; all figures show electron density contoured at 1.0σ.

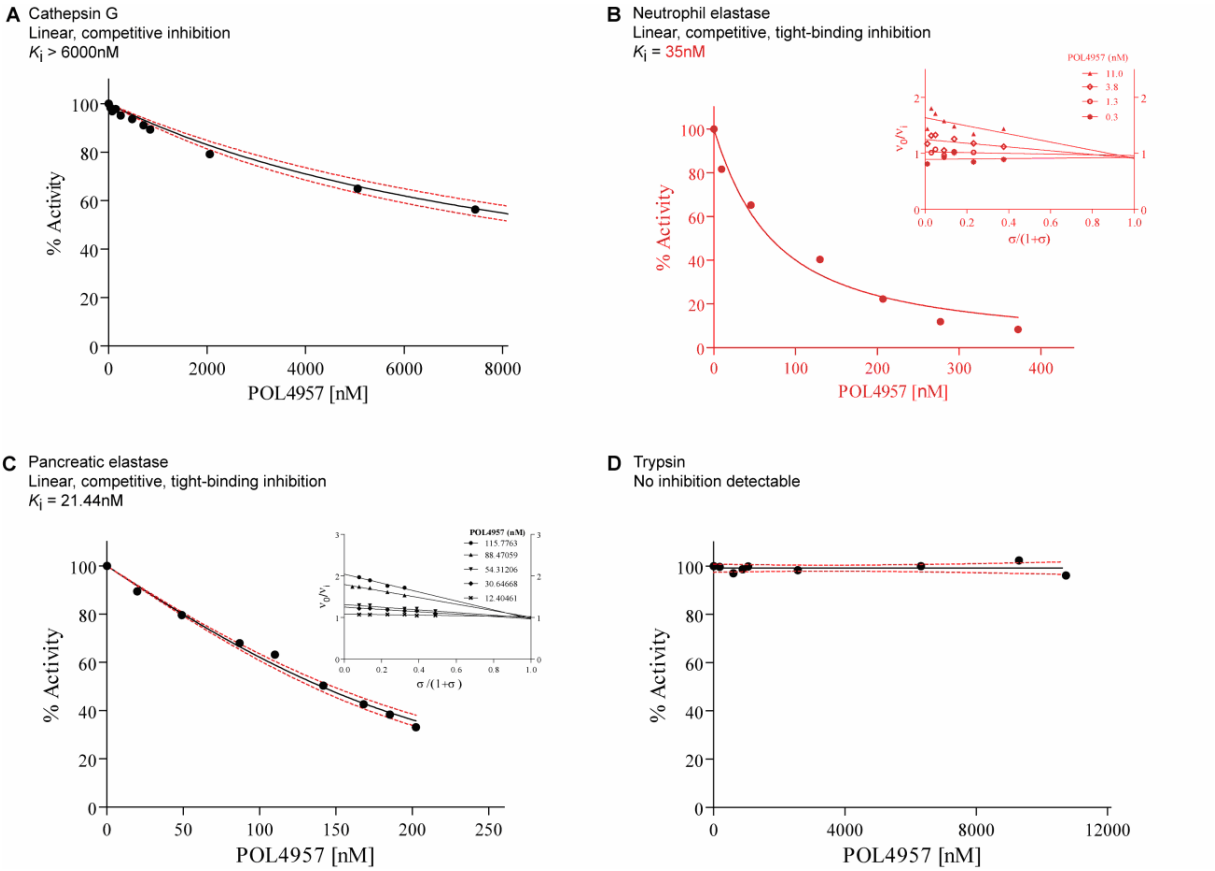
Figure 5)



Preliminary inhibitory data

Inhibition profile of POL4634 for cathepsin G, neutrophil elastase, pancreatic elastase and trypsin. Data were obtained fluorometrically with λex/λem = 360/465 nm. Best-fit kinetic parameters are shown. A) cathepsin G (227 nM) and the substrate Suc-Ala-Ala-Pro-Phe↓AMC at pH 6.5; [S] = 715 uM, K_m = 667 uM. B) Neutrophil elastase (11 nM) and the substrate Suc-(OMe)-Ala-Ala-Pro-Val↓AMC at pH 7.5; [S] = 715 uM, K_m = 820 uM. C) Pancreatic elastase (195 nM) and the substrate Suc-(OMe)-Ala-Ala-Pro-Val↓AMC, at pH 7.5; [S] = 996 uM, K_m = 231 uM. D) Trypsin (1.75 nM) and the substrate Boc-Gln-Ala-Arg↓AMC, at pH 7.5; [S] = 140 uM, K_m = 10 uM. The inset shows the specific velocity plot at five substrate and five inhibitor concentrations.

Figure 6)



Preliminary inhibitory data

Inhibition profile of POL4957 for cathepsin G, neutrophil elastase, pancreatic elastase and trypsin. Data were obtained fluorometrically with λex/λem = 360/465 nm. Best-fit kinetic parameters are shown. A) cathepsin G (227 nM) and the substrate Suc-Ala-Ala-Pro-Phe↓AMC at pH 6.5; [S] = 715 uM, K_m = 667 uM. B) Neutrophil elastase (11 nM) and the substrate Suc-(OMe)-Ala-Ala-Pro-Val↓AMC at pH 7.5; [S] = 715 uM, K_m = 820 uM. C) Pancreatic elastase (195 nM) and the substrate Suc-(OMe)-Ala-Ala-Pro-Val↓AMC, at pH 7.5; [S] = 996 uM, K_m = 231 uM. D) Trypsin (1.75 nM) and the substrate Boc-Gln-Ala-Arg↓AMC, at pH 7.5; [S] = 140 uM, K_m = 10 uM. The inset shows the specific velocity plot at five substrate and five inhibitor concentrations.

Figure 7A)

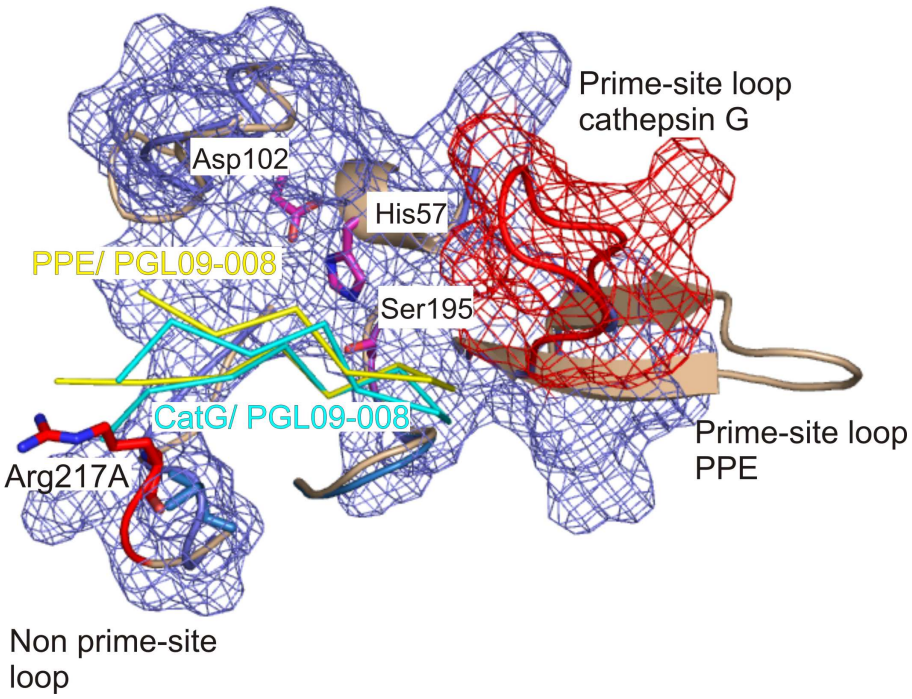


Figure 7B)

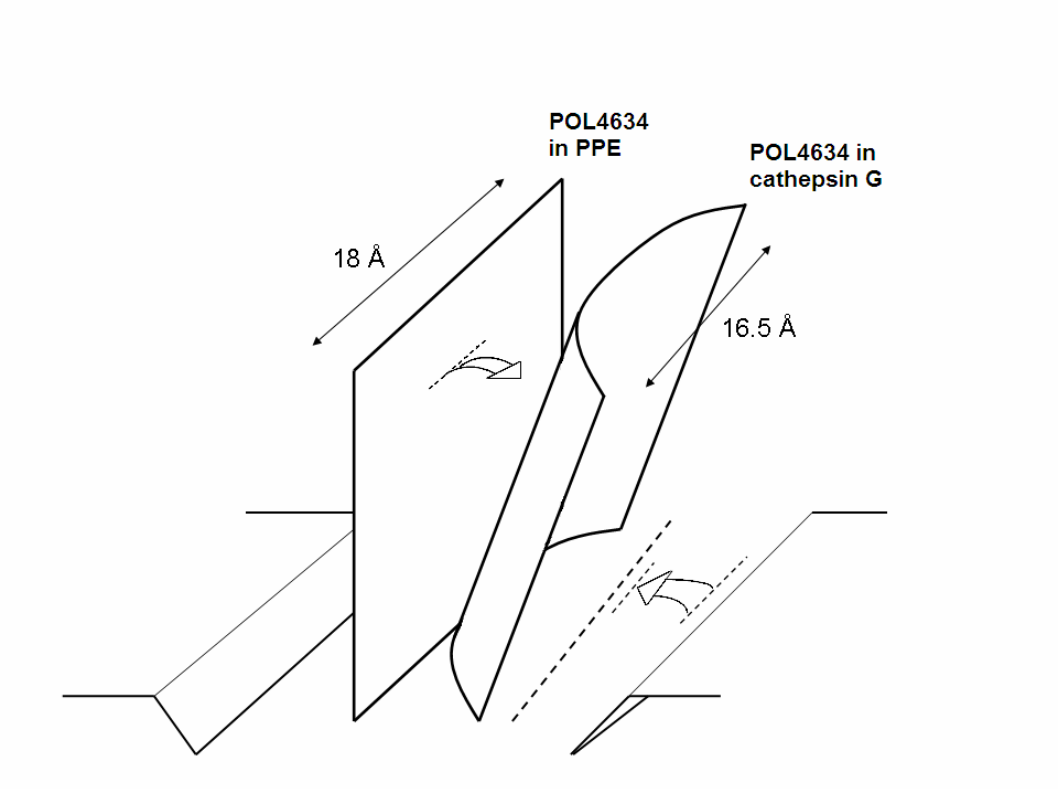


Figure 7)

A) Superposition of the active sites of cathepsin G and PPE with POL4634. The active site groove of cathepsin G is shown as cartoon in dark blue covered with a surface mesh. PPE is shown as cartoon coloured in raspberry. The prime-site loop of cathepsin G, shown in red, is perpendicularly oriented relative to the prime-site loop of PPE and the active site groove. Arg217A in the non prime-site loop of PPE is shown as sticks in red. POL4634 is shown as ribbon in light blue and yellow, respectively.
B) Schematic drawing of the POL4634 inhibitor in the active sites of cathepsin G and PPE, respectively.

Materials and methods

Inhibitors

The inhibitor POL4634 was provided by Polyphor, Allschwil Switzerland.

Substrates

All substrates were purchased from Peptanova, Sandhausen, Germany.

Enzymes

Human neutrophil cathepsin G was purchased from Lee Biosolutions, INC. St. Louis, USA. Porcine pancreatic elastase was purchased from Worthington Biochemical Corporation, Lakewood, USA. Bovine trypsin was purchased from Roche, Switzerland.

Enzyme Solutions

Human cathepsin G, porcine pancreatic elastase and bovine trypsin were dissolved in 50mM sodium acetate, pH 5.5 containing 150mM NaCl to a final concentration of 4 mg per ml. Small aliquots of this stock solution were stored at -80°C until used.

For cathepsin G, 10µl of stock solution were diluted in 2000µl AHNP-buffer (150mM sodium acetate, pH 6.5, 150mM HEPES, 300mM NaCl, 0.001% Pluronic). For porcine pancreatic elastase and bovine trypsin 10µl stock of solution were diluted in 2000µl assay buffer (50mM Tris-HCl, pH 7.5, 100mM NaCl, 0.01% Triton).

General Assay Conditions

The kinetic mechanism of the inhibition by POL4634 was studied using the substrates Suc-Ala-Ala-Pro-Phe↓AMC, Suc-(OMe) Ala-Ala-Pro-Val↓AMC and Boc-Gln-Ala-Arg↓AMC (all Peptanova, Sandhausen, Germany) for human cathepsin G, porcine pancreatic elastase and bovine trypsin respectively. All kinetic experiments were carried out in black 96-well Greiner CellStar plates in a total reaction volume of 100 µl and substrate hydrolysis was followed flurometrically at $\lambda_{ex}/\lambda_{em}$ = 360/465 nm using a Tecan Genious™ plate reader.

Determination of K_m

In parallel to each inhibition assay, the *Michaelis* constant (K_m) was determined under the assay conditions. The linear increase of the fluorescence was monitored at 12 different substrate concentrations (0 – 1600 µM). The Michaelis-Menten equation (**Equation 1**) was fitted to the steady-state (v) rates to calculate K_m by nonlinear regression.

$$v = \frac{V[S]}{K_m + [S]} \qquad \text{Equation 1}$$

Active Site Titration of Porcine Elastase

The concentration of the active sites in the PPE samples were determined flurometrically by titration with the irreversible inhibitor Suc-(OMe)-Ala-Ala-Pro-Val-CH_2Cl. After incubation of 30 min at room temperature the reaction was assayed. The residual activity was plotted against the amount of Suc-(OMe)-Ala-Ala-Pro-Val-CH_2Cl and the concentration of PPE was determined by means of a linear fit of the data points. The procedure was repeated twice to confirm its reproducibility.

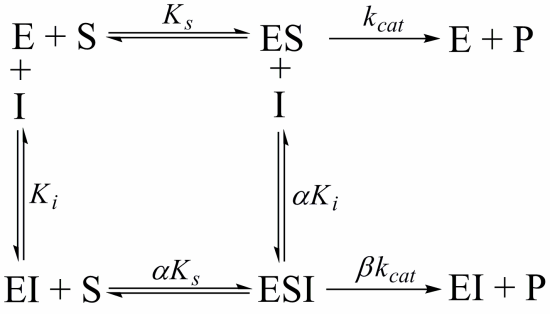
Inhibition Assay

Inhibitor and substrate solutions were premixed to a volume of 50 µl and the reaction was started with 50 µl enzyme solutions to a total reaction volume of 100 µl. DMSO concentration in the assay was kept constant throughout one entire measurement.

Data Analysis and Assignment of the Inhibition Mechanism

The kinetic mechanism of pancreatic Elastase was investigated in two steps using Suc-(OMe)-Ala-Ala-Pro-Val↓AMC at pH 7.5. In a first step the inhibition type was diagnosed using a combination of five substrate concentrations (0.25 to 5 times K_m), and five inhibitor concentrations in the range 0.1 to 6.7 µM. Analysis of initial velocity data for the 25 combinations of substrate and inhibitor concentrations was performed with the specific velocity plot, a graphical method suitable for diagnosing linear and hyperbolic inhibition mechanisms (Baici, 1981). This method is based on the general modifier mechanism shown in Scheme 1. Here, E represents the free enzyme, S the substrate, I the inhibitor, and P

collectively the products of the reaction. Other symbols are as follows: K_i = [E][I]/[EI], the competitive inhibition constant, αK_i = [ES][I]/[ESI], the uncompetitive inhibition constant, K_s = [E][S]/[ES], the substrate dissociation constant; αK_s = [EI][S]/[ESI]; k_{cat}, the catalytic constant. The dimensionless coefficients α and β characterize the proportion of competitive and uncompetitive character in the inhibition mechanism (Scheme 1)



Scheme 1

The specific velocity equation is as follows:

$$\frac{v_o}{v_i} = \frac{[I]\left(\frac{1}{\alpha K_i} - \frac{1}{K_i}\right)}{1+\beta\frac{[I]}{\alpha K_i}}\frac{\sigma}{1+\sigma} + \frac{1+\frac{[I]}{K_i}}{1+\beta\frac{[I]}{\alpha K_i}}$$

Equation 2

where v_0 and v_i are the initial velocities in the absence and in the presence of inhibitor, respectively, σ = [S]/K_m, and other symbols as described above. From the specific velocity plot and its secondary plots, sufficiently precise values of α and β could be calculated. In a second step, an accurate value of K_i was obtained by fitting equation 2 (Szedlacsek et al., 1988) to initial velocity data measured at fixed enzyme and substrate concentrations and several POL4634 concentrations (Figures 5 and 6) by nonlinear regression using GraphPad Prism version 5 for Windows, GraphPad Software, San Diego, California USA (www.graphpad.com). The variables v_0 and v_i are the initial velocities in the absence and in the presence of inhibitor, respectively, $[E]_t$ is the total enzyme concentration, $[I]_t$ is the total inhibitor concentration, σ = [S]/K_m and α and β are explained above.

$$v_i = \frac{v_0}{2}\left[\frac{\alpha+\sigma-\beta(1+\sigma)}{\alpha+\sigma}\right]\left\{\sqrt{\left[\left(\frac{1+\sigma}{\alpha+\sigma}\frac{\alpha K_i}{[E]_t}+\frac{[I]_t}{[E]_t}-1\right)^2+4\frac{1+\sigma}{\alpha+\sigma}\frac{\alpha K_i}{[E]_t}\right]}+\frac{\alpha+\sigma+\beta(1+\sigma)}{\alpha+\sigma-\beta(1+\sigma)}-\frac{1+\sigma}{\alpha+\sigma}\frac{\alpha K_i}{[E]_t}-\frac{[I]_t}{[E]_t}\right\} \quad \text{Equation 3}$$

For cathepsin G initial velocity data were directly fitted to equation 2, with $\alpha = \infty$ and $\beta = 0$ for linear competitive inhibition.

Crystallisation

PPE was purchased from Worthington Elastase (New Jersey, USA) and cathepsin G from Elastin Products (Missouri, USA) and crystallized without further purification. PPE was dissolved in 10 mM sodium acetate, pH 5.0, and cathepsin G was dissolved in 50 mM ammonium sulphate, 50 mM Tris, pH 8.0 at final concentrations of 0.8 mM. The proteases were incubated with two-fold molar excess of POL4634 (dissolved in 4 % DMF/water). Due to the limited water solubility of POL4634 the protease/inhibitor mixtures were turbid at the beginning of the incubation period, but turned clear after one to two hours incubation at room temperature. Crystals were obtained using the vapor diffusion method with 96-well sitting drop plates (Greiner bio-one, Germany) containing mixtures of 100 nl protein solution and 100 µl reservoir solution. Crystals of PPE/POL4634 complexes were obtained within 2 weeks at 17 % PEG 8000, 20 mM TAPS, pH 9.0. The cathepsin G/POL4634 complex was crystallized at 25 % PEG 8000, 150 mM ammonium sulphate, 50 mM Tris, pH 8.5. The crystals were flash frozen in liquid propane in the mother liquor supplemented with 20 % glycerol as a cryo-protectant.

Structure solution and refinement

The X-ray data was collected at 100 K at the Swiss synchrotron light source (Villigen, Switzerland) at a wavelength of 1.0 Å. Data were processed with the program XDS [Kabsch, 1993]. Initial phases were obtained by molecular replacement using the program PHASER [McCoy, 2007] using known PPE (PDB-ID: 3EST) and cathepsin G (PDB-ID: 1T32) structures as search templates. The three-dimensional structures were refined using

REFMAC5 [CCP4] and SHELXL [Sheldrick, 1997], and the model building was performed using the program COOT [Emsley, 2004]. Restraints for the non-natural amino acids ^{D}Pro and NLE were generated using the PRODRG server [Schuettelkopf, 2004]. All figures were made using Pymol [DeLano, 2002].

Structure analysis

The shape complementarity (SC) was calculated using SC [CCP4]. In order to calculate the SC, D-proline was renamed into PRO, octyl-glycine and norleucine were modified into leucines and all double conformations were eliminated. Hydrogen bonds were analyzed with HBPLUS [McDonald, 1993] and the interface was calculated using the PISA server v1.18 [Krissinel, 2007].

Acknowledgements

We gratefully acknowledge Antonio Baici and Patricia Schenker for assistance of the evaluation of kinetic data.

References

1. Luckett, S., Santiago Garcia, R., Barker, J.J., Konarev, AI.V., Shewry, P.R., Clarke, A.R., Brady, R.L.,, *High-resolution structure of a potent, cyclic proteinase inhibitor form sunflower seeds.* J Mol Biol, 1999. **290**(2): p. 525-33.

2. Daly, N.L., et al., *The absolute structural requirement for a proline in the P3'-position of Bowman-Birk protease inhibitors is surmounted in the minimized SFTI-1 scaffold.* J Biol Chem, 2006. **281**(33): p. 23668-75.

3. Garcia Boy, R., Mier, W., Nothelfer, E.M., Altmann, A., Eisenhut, M., Kolmar, H., Tomaszowski, M., Krämer, S., Haberkorn, U., *Sunflower Trypsin Inhibitor 1 Derivatives as Molecular Scaffolds for the Development of Novel Peptidic Radiopharmaceuticals.* Mol Imaging Biol, 2009.

4. Descours, A., et al., *A new family of beta-hairpin mimetics based on a trypsin inhibitor from sunflower seeds.* Chembiochem, 2002. **3**(4): p. 318-23.

5. Legowska, A., et al., *Introduction of non-natural amino acid residues into the substrate-specific P1 position of trypsin inhibitor SFTI-1 yields potent chymotrypsin and cathepsin G inhibitors.* Bioorg Med Chem, 2009. **17**(9): p. 3302-7.

6. Li, P., Jiang, S., Lee, S.L., Lin, C.Y., Johnson, M.D., Dickson, R.B., Michejda, C.J., Roller, P.P., *Design and Synthesis of Novel and Potent Inhibitors of the Type II Transmembrane Serine Protease Matriptase, Based upon the Sunflower Trypsin Inhibitor-1.* J Med Chem, 2007. **50**(24): p. 5976-83.

7. Makhatadze, G.I., Kim, K-S., Woodward, C., Privalov, P.L., , *Thermodynamics of BPTI folding.* Protein Science, 1993. **2**(12): p. 2028-36.

8. Colgrave, M.L., Craik, D.J.,, *Thermal, chemical and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot.* biochemistry, 2004. **43**(20): p. 5965-75.

9. Qi, R.F., Song, Z.W., Chi, C.W.,, *Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application.* Acta Biochim Biophys Sin (Shanghai), 2005. **37**(5): p. 283-92.

10. Avrutina, O., Schmoldt, H.U., Gabrijelcic-Geiger, D., Le Nguyen, D., Sommerhoff, C.P., Diederichsen, U., Kolmar, H.,, *Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTi-II.* Biol Chem, 2005. **386**(12): p. 1301-06.

11. Craik, D.J., Daly, N.L., Bond, T., Waine, C.,, *Plant cyclotides: A unique family of cyclic knotted proteins that defines the cyclic cysteine knot structural motif.* J Mol Biol, 1999. **294**(5): p. 1327-36.

12. Laskowsi, M.a.Q., M.A.,, *What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?* Biochim Biophys Acta, 2000. **1477**(1-2): p. 324-37.

13. Apostoluk, W., Otlewski, J.,, *Variability of the canonical loop conformations in serine proteinases inhibitors and other proteins.* Proteins, 1998. **1998**(32): p. 4.

14. Tyndall, J.D., Nall, T., Fairlie, D.P.,, *Proteases universally recognize beta-strands in their active sites.* Chem Rev., 2005. **105**(3): p. 973-99.

15. Hansen, D., Macedo-Ribeiro, S., Verissimo, P., Yoo Im, S., Sampaio, M.U., Oliva, M.L.,, *Crystal structure of a novel cysteineless plant Kunitz-type protease inhibitor.* Biochem Biophys Res Commun, 2007. **360**(4): p. 735-40.

16. Garavilla de, L., Greco, M.N., Sukumar, N., Chen, Z-W., Pineda, A.O., Mathews, F.S., Cera di, E., Giardino, E.C., Wells, G.I., Haertlein, B.J., Kauffman, J.A., Corcoran, T.W., Derian, C.K., Eckardt, A.J., Damiano, B.P., Andrade-Gordon, P., Maryanoff, B.E, *A novel, potent dual inhibitor of the leukocyte proteases cathepsin G and chymase.* J Biol Chem, 2005. **280**(18): p. 18001-7.

17. Zani, M.-L., Baranger, K., Guyot, N., Dallet-Choisy, S., Moreau, T., *Protease Inhibitors Derived from Elafin and SLPI and Engineered to Have Enhanced Specificity Towards Neutrophil Serine Proteases.* Protein Science, 2009. **18**: p. 579-594.

18. Legowska, A., et al., *Selection of peptomeric inhibitors of bovine alpha-chymotrypsin and cathepsin G based on trypsin inhibitor SFTI-1 using a combinatorial chemistry approach.* Mol Divers, 2010. **14**(1): p. 51-8.

19. Hilpert, K., Hansen, G., Wessner, H., Volkmer-Engert, R., Höhne, W.,, *Complete substitutional analysis of a sunflower trypsin inhibitor with different serine proteases.* J Biochem, 2005. **138**: p. 383-390.

20. Thongyoo, P., Bonomelli, C., Leatherbarrow, R.J., Tate, E.W., *Potent Inhibitors of beta-Tryptase and Human Leukocyte Elastase Based on the MCoTI-II Scaffold.* J Med Chem Letters, 2009. **52**: p. 6197-6200.

21. Brillard-Bourdet, M., Hamdaoui, A., Hajjar, E., Boudier, C., Reuter, N., Ehret-Sabatier, L., Bieth, J.G., Gauthier, F., *A Novel (Schistocerca gregaria) Serine Protease Inhibitor with a High Affinity for Neutrophil Elastase.* Biochem J, 2006. **400**: p. 467-476.

22. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants.* Journal of Applied Crystallography, 1993. **26**: p. 795-800.

23. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, C.C., Read, R.J., *PHASER.* J. Appl. Crystallogr., 2007. **40**: p. 658-74.

24. CCP4, *Collaborative Computational Project No. 4.* Acta Cryst., 1994. **D**(50): p. 760-63.

25. Sheldrick, G.M., Schneider, T.R.,, *SHELXL: high-resolution refinement* Methods Enzymol, 1997. **277**(319-43).

26. Emsley, P., Cowtan, K.,, *Coot: model-building tools for molecular graphics.* Acta Cryst D, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

27. Schuettelkopf, A.W., Aalten van, D.M.F.,, *PRODRG - a tool for high-throuput crystallography of protein ligands and complexes.* Acta Cryst D, 2004. **60**: p. 1355-63.

28. DeLano, W.L., *The pymol molecular graphics system*. 2002, DeLano Scientific: San Carlos, CA, USA.

29. CCP4, *Collaborative Computational Project No. 4.* Acta Cryst D, 1994. **50**(760-63).

30. McDonald, I., Naylor, D., Jones, D., Thornton, J.,, *HBPLUS Hydrogen bond calculator v. 3.15.* 1993.

31. Krissinel, E., Henrick, K.,, *Inference of macromolecular assemblies from crystalline state.* JMB, 2007. **372**: p. 774-97.

CHAPTER 3

STRUCTURAL CHARACTERIZATION OF A NTN BETA-AMINOHYDROLASE FROM SPHINGOSINICELLA XENOPEPTIDILYTICA

INTRODUCTION

The superfamily of *N*-terminal nucleophile aminohydrolases[1] is divided into seven families according to the structural classification of proteins database (SCOP): the class II glutamine amino-transferases, the penicillin acylase, the penicillin V acylase, the proteasome subunits, the (glycosyl)-asparaginase, the γ-glutamyl trans-peptidase-like and the SPO2555-like families. The enzymes all share a common αββα-sandwich-core fold, but have their unique substrate preferences. The enzymes either posses an amino-terminal serine, threonine or cysteine residue that act as primary nucleophile.

The heterooctameric structure of DmpA[2] from *Ochrobactrum anthropi* has been solved by Bompard-Gilles et al. [1]. Each of the four αβ-subunits forms the αββα-sandwich fold and exposes an active site serine residue, which is situated at the amino-terminus of the β-subunit. The enzyme undergoes self-processing. The SCOP database currently classifies DmpA in its own superfamily named DmpA/ArgJ-like DmpA-like refers to the family name and DmpA/ArgJ to the name of its fold.

BapA[3] from *Sphingosinicella xenopeptidilytica* 3-2W4 shares 42.2% sequence identity with DmpA. Both enzymes are structurally and functionally very similar. Originally, BapA was discovered by growing a heterogeneous mixture of microbial cultures taken from a waste water treatment plant in β-dipeptide and β-tripeptide supplemented minimal media. After several enrichment cycles, the bacterial strain 3-2W4 was isolated and later characterized by DNA sequencing as "Sphingomonas" strain. At this stage, the newly discovered enzyme was not characterized, but it could be tracked and isolated without difficulty by measuring its cleavage activity of *p*-nitroanilide coupled beta-amino acids during multi-step purification procedures. Its presence was detected by measuring the absorbance of free *p*-nitroanilide at 405 nm. Finally, enriched protein fractions were digested and sequenced. A database search with the *N*-terminal sequence fragment revealed a so far unknown enzyme, subsequently named BapA, according to its β-amino-peptidase activity. Native BapA, similarly to DmpA, undergoes self-proteolytic cleavage and exposes an amino-terminal active site serine residue, which is situated at the amino-terminus of the β-subunit. Two subunits each form a compact αββα-sandwich fold that is characteristic for Ntn hydrolases [2].

Both hydrolases, BapA and DmpA, exhibit rather unusual peptidase, esterase and amidase specificities. BapA was first characterized by Geueke et al. in 2005. The authors show the unusual substrate specificity of BapA for substrates with β-amino acids at the amino-terminus with highest preference for aliphatic and aromatic residues [2], [3]. Interestingly, peptide substrates with *N*-terminal α-amino acids are not cleaved. DmpA, however, also accepts beta-amino-acid substrates, but the *N*-terminal residue is limited to either a beta-glycine or beta-alanine. Furthermore, DmpA accepts substrates with *N*-terminal alpha-amino acids with both D- and L-configuration [1].

[1] *N*(amino) terminal *n*(amino) hydrolase; Ntn
[2] L-aminopeptidase-D-amidase/D-esterase; DmpA
[3] Beta-peptidyl aminopeptidase (hydrolase); BapA

Table 1_3) annotated protein database (PDB) files of Ntn hydrolases

	PDB ID	Protein and redundant PDB entry	Active form	Pro-form	spacer	SCOP
1	1B65	L-aminopeptidase DmpA	(αβ)4	(α)4	no	X
2	2PVA	penicillin V acylase 2Z71, 2QUY, 1GM7, 2OQC (*B. subtilis*), 3PVA (*B. sphaericus*)	(α)4	nd.	no	3
3	2BJF 2BJG	bile acid hydrolase with taurine and deoxycholate	(α)4	nd.	no	0
4	2GL9	Glycosylasparaginase	(αβ)2	nd.	nd.	0
5	1E3A 1GK9	penicillin G acylase (PA) (*E. coli*) PA with substrate 1FXH, 1FXV, 1JX9, 1K5Q	(αβ)	(α)	yes	2
6	1OQZ	glutaryl-7-aminocephalosporanic acid acylase1GK0, 1OR0	(αβ)2	(α)	yes	2
7	2NQ0	γ-glutamyltranspeptidase (*Helicobacter pylori*) 2QM6, 2QMC (2QM6 with glutamate)	(αβ)	(α)	no	6
8	2V4I	glutamate N-acetyl transferase 2 2VZK	nd.	nd.	nd.	0
9	1KUU	MTH1020 from *Methanobacterium thermoautotrophicum*	nd.	nd.	nd.	0
10	2NTK	PurO (*Methanothermobacter Thermoautotrophicus*)	α	nd.	no	0
11	1GK0	glutarylamidase (*Pseudomonas sp.*) glutarylamidase (*Pseudomonas diminuta*)	(αβ)2 (αβ)	(α)2 (α)	Yes yes	2
12	1JN9	putative asparaginase, ybik gene (*E. coli*) 1KZX (2ZAL with aspartate)	(αβ)	nd.	no	5
13	2E0W	γ-glutamyltranspeptidase (*E. coli*)	(αβ)	(α)	no	6
14	1G0U	proteasome component of 20S 1G65, 2GPI, 2GPL				4
15	2ZAK	isoaspartyl peptidase / L-asparaginase Ec AIII (3C17)	(αβ)	(α)	no	0
16	2GEZ	plant asparaginase	(αβ)	(α)	no	0
17	1APY 1APZ	human aspartylglucosaminidase + reaction product	(αβ)2	(α) / (α)2	no	5

nd: not determined or indicated in literature; SCOP classification: 1: Class II glutamate amidotransferase; 2: Penicillin acylase, catalytic domain; 3: Penicillin V acylase; 4: Proteasome subunits; 5: (Glycosyl) asparaginase; 6: Gamma glutamyltranspeptidase-like; 7: SPO2555-like; 0: not annotated in SCOP data-base; X: no Ntn hydrolase according to SCOP.

1. Ntn hydrolases

17 different proteins (Table 1_3, 49 proteins in total) of the class of Ntn hydrolases are currently annotated in the protein databank.

The first notion of Ntn hydrolases dates back to the year 1995, when the crystal structures of glutamine PRPP amidotransferase, penicillin acylase and the proteasome component 20S were determined [4], [5], [6]. Those enzymes share the same fold and use the side chain of an amino-terminal residue of a β-sheet as a nucleophile for their catalytic activity. At that time, Brannigan et al. were the first to suggest the name Ntn (*N*-terminal nucleophile) hydrolase and the classification of the aforementioned proteins into this superfamily [7].

1.1 The Ntn fold

The core of the Ntn hydrolase consensus fold comprises two layers of five antiparallel β-sheets that are flanked by two α-helices (schematically illustrated in Figure 1_3). The overall fold is known as αββα-sandwich.

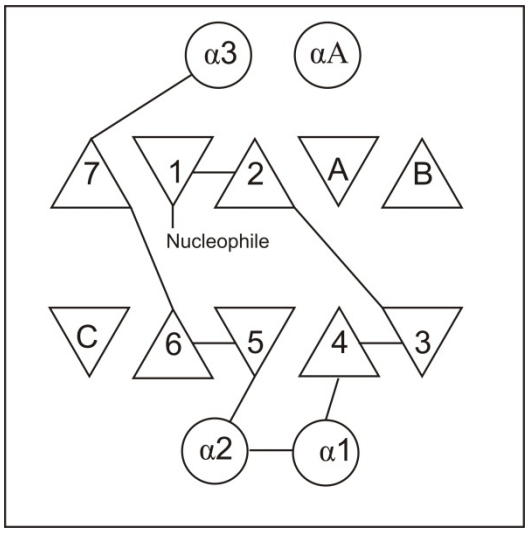


Figure 1_3)
Consensus fold of Ntn hydrolases: triangles represent β-sheets, circles represent α-helices. The αββα-sandwich fold comprises two layers of five anti-parallel β-sheets that are flanked by two α-helices. The numbered secondary structure elements occur in the same position and sequence in penicillin acylase, glutamine 5-phosphoribosyl-1-pyrophosphate amido-transferase, glycosyl asparaginase and proteasome subunits. The β-sheets A, B and C and the helix αA are structurally conserved but not always in the same sequence order. The figure is adapted and modified from Bompard-Gilles et al, 2000 [1].

It is important to note that the connectivity and the number of additional secondary structure elements differ from protein to protein. Penicillin G acylase for example comprises additional α-helices and β-sheets and its Ntn fold consists only of a single

chain (β-chain), in contrast to DmpA, which is composed of two chains (α- and β-chain) upon self-processing. Thus, the consensus fold depicted in Figure 1_3 is a simplified view and not necessarily correct in particular cases. However, it facilitates protein family classifications displaying this consensus fold.

1.2 Multimeric states of Ntn hydrolases

Ntn hydrolases exist in diverse multimeric states. The inactive monomeric, dimeric or tetrameric single precursor polypeptide chains are autoproteolytically processed into (αβ) heterodimers, tetramers or octamers, respectively. Studies of Ntn hydrolases on the assembly formation in the context of the self-processing have not been reported so far, hence, it is not known whether the dimeric or tetrameric inactive quaternary structures of the precursor must form prior to the autoproteolytic cleavage of every subunit, or whether monomeric species undergo the autoproteolytic cleavage prior to the quaternary assembly formation. Some members of the Ntn superfamily, however, are active as single polypeptide chains (see Table 1_3). The only known reported exception is glutaryl-7-aminocephalosporin acylase, which is monomeric in its pro-form and dimerizes into a $(\alpha\beta)_2$ heterotetramer upon activation [8]. Penicillin V acylase and bile acid hydrolase are homotetramers [9], [10]. They are active upon leader peptide cleavage and do not require further self-processing reactions. Thus, the precursor processing for these two enzymes already results in active Ntn hydrolases (Table 1_3).

1.3 Pro-form activation

The superfamily of Ntn hydrolases is highly heterogeneous with respect to polypeptide processing and oligomerization. Typically, Ntn hydrolases are synthesized as pre-pro-polypeptide chains including an *N*-terminal periplasmic pre-leader sequence. Penicillin acylase from *E. coli* is the only known example that requires the pre-sequence for enhanced and correct folding [8]. Other Ntn hydrolases are expressed as pro-polypeptide precursors which undergo a single autoproteolytic cleavage reaction. It is not known, whether the maturation process into subunits of heterodimers is a consequence of the formation of multimeric states before self-processing or whether the assembly formation is a prerequisite for the self-processing. However, cephalosporin acylase and glutaryl-7-aminocephalosporanic acid acylase are further processed by a second intermolecular cleavage reaction releasing a short internal spacer sequence [8], [11], [12].

Recently, Michalska et al. published a general mechanism for the pro-form activation on the structural basis of a plant-type L-asparaginase [13]. It has been proposed that the first common step of the auto-proteolytic cleavage is an *N→O* (serine or threonine) or *N→S* (cysteine) acyl shift through the formation of an oxyoxazolidine or oxythiazolidine intermediate. Hereby, the "acyl-group" represented by the *C*-terminal end (peptide bond) is transferred from the "*N*-terminus" onto the oxygen or sulphur atom of the side chain, temporarily featuring an ester or thioester bond. Correspondingly, the first half-reaction is called acyl shift. This rearrangement is achieved via the deprotonation of the active site

residue by a not further specified "general base" that can either be a hydrogen bonded water molecule or a basic residue. The deprotonated active site residue then attacks the carbonyl carbon of the peptide bond *N*-terminal of the active site residue. The oxyoxazolidine or oxythiazolidine transition state is stabilized by the oxyanion hole formed by Asn67 and a water molecule. Structural rearrangements lead to the formation of the (thio)-ester acyl enzyme intermediate. Then, the ester bond is hydrolyzed and the carboxy-terminal end is released. This pathway is illustrated in Figure 2_3. A number of distorted high-energy trans-peptide bonds in aspartylglucosaminidase, glutaryl-7-aminocephalosporanic acid acylase and cephalosporin acylase have been reported that are made responsible for the efficient displacement of the carboxy-terminal tail of the α-subunit form the active site cleft [8], [12], [14].

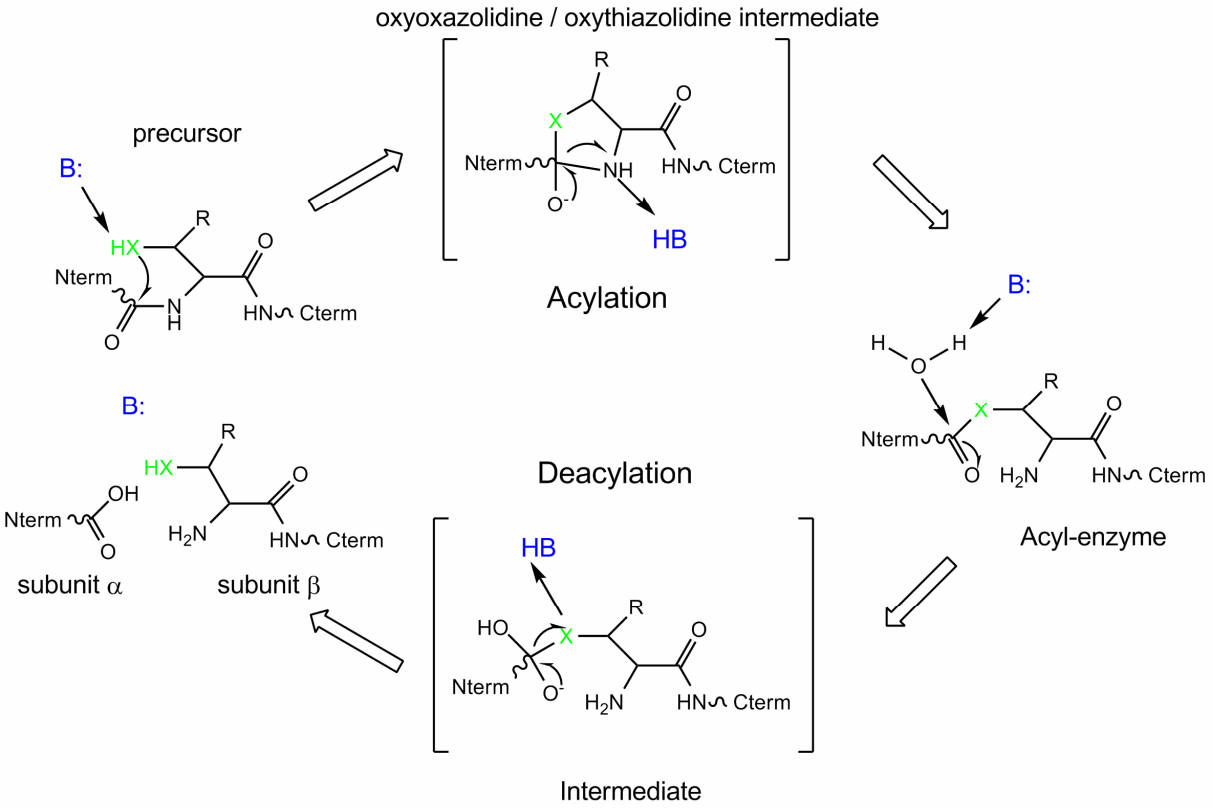


Figure 2_3)
Proposed autoproteolysis mechanism of an L-asparaginase. The oxyoxazolidine intermediate state is formed by the active site serine or threonine residue; cysteine forms the oxythiazolidine intermediate state. The nucleophilic "X" denoted in green either represents sulphur (cysteine) or oxygen (serine / threonine), "B" denoted in blue is the general base; R = H (Ser), R = CH_3 (Thr). The figure is adapted and modified from [13].

1.4 Catalytic mechanism

In contrast to the autoproteolytic pro-form activation, a consensus mechanism for the catalytic activity of mature Ntn hydrolases has not been proposed in the literature so far. This might be due to the different nature and chemistry of the substrates and the number

of available structures with a substrate, substrate analogue or inhibitor bound. Currently, five such complex structures can be retrieved from the protein database (bile acid hydrolase, penicillin G acylase, γ-glutamyltranspeptidase, putative asparaginase and human aspartylglucosaminidase) (see Table 1_3) and all these Ntn hydrolases exhibit different functions, making it difficult to propose a general mechanism.

1.5 Beta-aminohydrolase A

BapA features two interesting properties. Firstly, it binds penicillin based antibiotics, in particular ampicillin and β-lactamase hydrolyzed ampicillin, and secondly, BapA exhibits reversible beta-peptidyl aminohydrolase activity. Thus, BapA is also able to perform β-peptide synthesis [22]. Since its discovery, BapA has been thoroughly characterized with respect to its substrate specificities [3] and it has been investigated for industrial synthesis of short beta-peptides. Currently, the most interesting synthesis compound candidate is L-carnosine (β-Gly-α-His-dipeptide) (Figure 3_3). L-carnosine has a positive effect on the repair of injured intestinal epithelial cells [15], a down-regulating effect on the release of pro-inflammatory cytokines [16] and anti-ageing effects [17]. Therefore, circumventing chemical syntheses which require organic solvents and heavy metal catalysts would be highly desired to be able to supply the cosmetics market for L-carnosine.

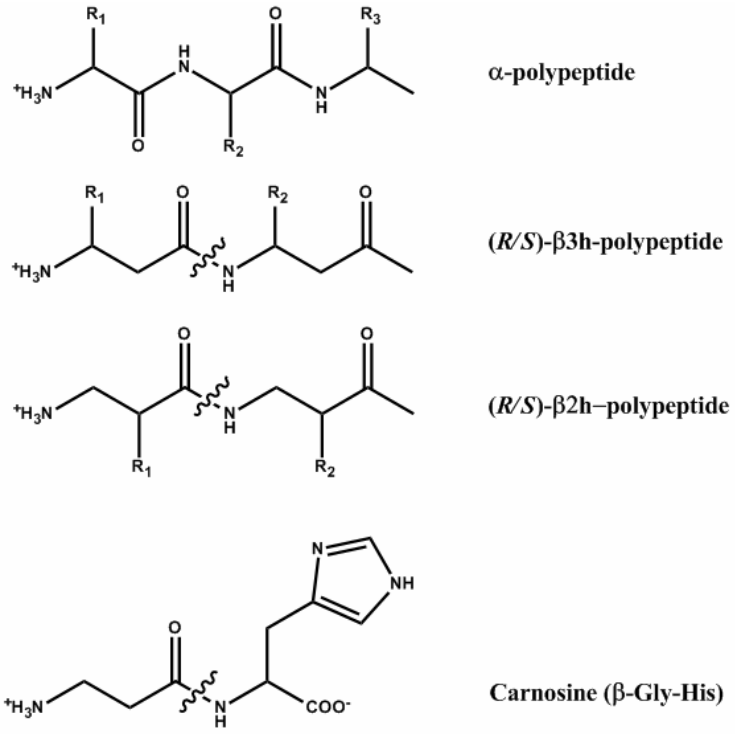


Figure 3_3)
Schematic representation of alpha- and beta-amino acid polypeptides. Side chains are denoted with R1, R2 and R3. The wavy line indicates the putative BapA peptide cleavage site.

2. Rationale and aim of the project

The aim of the project is to solve the crystal structures of BapA, its non-cleaved pro-form and additionally various mutants to elucidate the essential residues for the self-processing, the substrate cleavage as well as the residues that determine the β-amino acid substrate specificity.

This project was carried out in collaboration with the EAWAG, Dübendorf, Switzerland (group of Dr. H-P. Kohler) and the ETH, Zürich, Switzerland (group of Prof. Dr. D. Seebach).

MATERIALS AND METHODS

Protein production and purification

The production and purification of BapA from *Sphingosinicella xenopeptidilytica* is described in detail in [3], [18]. The functional characterization of beta-peptides is described in [18] and [22].

Crystallization of proBapA

The active site serine (S250) was mutated to alanine to prevent autoproteolytic activation. This form will subsequently be named proBapA, implying the S250A mutant. ProBapA was crystallized in 24-well sitting drop plates (HR3-160 Cryschem, Aliso Viejo, CA, USA) using the vapour diffusion method at a protein concentration of 13.5 mg/ml in 2 μl drops (1 μl protein, 1 μl reservoir) in 1.0 – 1.2 M ammonium sulphate buffered at pH 7.0 to 7.5 (0.1 M HEPES (2-[4-(2-hydroxyethyl)piperazin-1yl]ethanesulfonic acid) at 4°C.

Data collection and processing of proBapA

ProBapA crystals were cryo-protected with 30% glycerol, flash frozen in liquid nitrogen and measured at the Swiss synchrotron light source (PSI Villigen, Switzerland). Diffraction data up to 1.8 Å resolution on a total of 720 frames were recorded with an oscillation range of 0.5° per frame. The data were processed with the program XDS [19]. Initial phases were obtained by molecular replacement using a truncated model of tetrameric BapA (residues 1-230 and 250-371) as search model with the program PHASER [20]. The structure (1 molecule / asymmetric unit) was refined to final R-factors of R_{work} = 17.4% and R_{free} = 19.4% in the space group P212121, using the program package PHENIX [21]. Statistics for diffraction data and structure refinement are summarized in Table 2_3.

Crystallization of BapA

Crystallization screenings for BapA were performed with a high throughput robot crystallization facility (Mycrosys 1000, Cartesian Technologies, Irvine CA, USA), on 96-well plates (Crystal Quick Plate, Greiner BioOne, Germany) at 20°C, stored and visualized by a Crystal Farm (Imaging Systems, Bruker Nonius; San Diego, CA, USA). Initial crystallization trials were performed with BapA at a concentration of 15 mg/ml at 20°C with mixing ratios of 50 nl protein / 150 nl reservoir, 100 nl protein / 100 nl reservoir and 150 nl protein / 50 nl reservoir. The experiment was performed in sitting drop plates allowing crystallization using the vapour diffusion method. The Clear Strategy ScreenTM I/II (MDL, Newmarket, Suffolk UK) in the pH range from 5.5 to 8.5 and the Sigma-Fluka Factorial Screen (#82009) were used. The first crystals were observed one weak after having set up the plates.

After several optimization trials, BapA at a concentration of 13 mg/ml was crystallized in 24-well HR3-160 Cryschem sitting drop plates (Aliso Viejo, CA, USA) in a temperature controlled room at 22°C in 30% - 50% ammonium sulphate (saturated solution at 4°C) at pH 7.0 – 8.5. The protein was dispensed in 2 μl drops (1:1 protein - reservoir ratio).

Data collection and processing of BapA

Crystals grown in 1.5 M ammonium sulphate, 100 mM HEPES pH 7.5, cryo-protected with 30% glycerol and flash frozen in liquid nitrogen were measured at the Swiss synchrotron light source (PSI Villigen, Switzerland). Diffraction data up to 1.45 Å resolution on a total of 720 frames were recorded with an oscillation range of 0.5° per frame. The data was processed with the program XDS [19]. Initial phases were obtained by molecular replacement using the program PHASER [20] and the structure of DmpA (protein database code 1B65) as search model. Figure 4_3 shows the electron density map after refinement. The asymmetric unit contains one molecule, which was refined to final R-factors of R_{work} = 15.0% and R_{free} = 16.8% in the space group P21 using the program package PHENIX [21]. Statistics for diffraction data and refinement are listed in Table 2_3.

Table 2_3) Statistics of data collection and refinement

Data	BapA	proBapA
Space group	P21 (SG 4)	P212121 (SG 19)
Cell parameters (Å)	a=87.4; b= 96.7, c= 101.4, β = 108.2	a=99.7; b= 113.7; c= 126.0
Resolution (Å)	1.45 (1.45-1.47)	1.8 (1.8-1.9)
Observed reflections	901804	579571
Completeness (%)[a]	98.5 (92.3)	99.1 (99.2)
Redundancy	3.23	4.4
Rsym (% on I)[a]	4.6 (43.3)	7.1 (47.3)
I/σI[a]	16.88 (2.68)	15.64 (3.96)
R / Rfree (%)	15.0 / 16.8	17.4 / 19.4
Ordered waters	1871	978
Bond lengths (Å)	0.004	0.012
Bond angles (°)	0.803	1.143
Average B factor (Å^2)	13.68	16.41
Res. most favoured regions	96.08 (1371)	96.49 (1346)
Res. generally allowed regions	3.43 (49)	2.8 (39)
Residues in disallowed regions	0.49 (7)	0.72 (10)

[a]: numbers in parenthesis indicate statistics in the last resolution shell

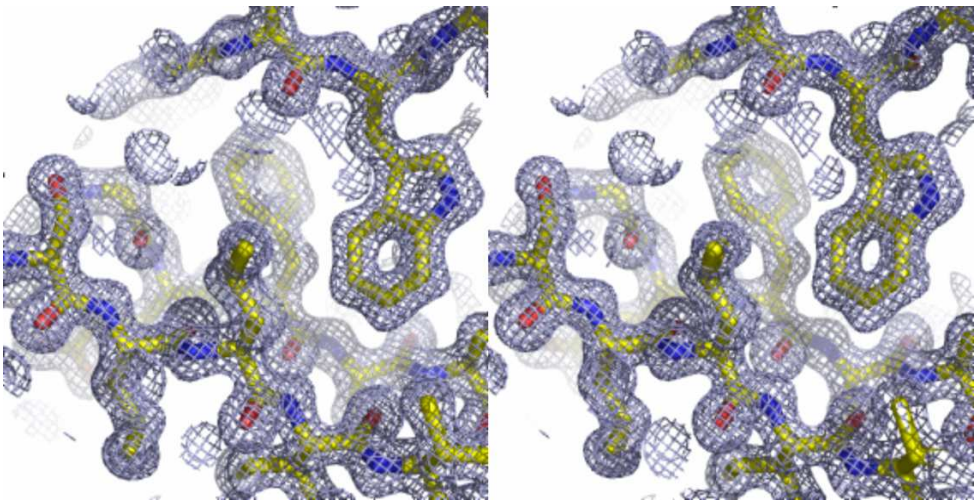


Figure 4_3)
2Fo-fc electron-density map of BapA contoured at 1.6σ, calculated using data at 1.45 Å resolution.

REFERENCES

1.) Bompard-Gilles, C., et al., A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-ala-esterase/amidase from Ochrobactrum anthropi. Structure, 2000. 8(2): p. 153-62.

2.) Geueke, B., et al., A novel beta-peptidyl aminopeptidase (BapA) from strain 3-2W4 cleaves peptide bonds of synthetic beta-tri- and beta-dipeptides. Journal of Bacteriology, 2005. 187(17): p. 5910-7.

3.) Geueke, B., et al., Bacterial beta-peptidyl aminopeptidases with unique substrate specificities for beta-oligopeptides and mixed beta,alpha-oligopeptides. FEBS Journal, 2006. 273(23): p. 5261-72.

4.) Smith, J.L., Zaluzec, E.J., Wery, J-P., Niu, L., Switzer, R.L., Zalkin, H. and Satow, Y, Structure of the Allosteric Regulatory Enzyme of Purine Biosynthesis. Science, 1994. 264: p. 1427-33.

5.) Duggleby, H.S., Tolley, S.P., Hill, C.P., Dodson, E.S., Dodson, G., Moody, P.C., Penicillin acylase has a single-amino-acid catalytic centre. Nature, 1995. 373(6511): p. 264-8.

6.) Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., Huber, R., Crystal Structure of the 20S Proteasome from the Archaeon T. acidophilum at 3.4 A Resolution. Science, 1995. 269: p. 533-539.

7.) Brannigan, J.A., Dodson, G., Duggleby, H.S., Moody, P.C., Smith, J.L., Tomchick, D.R., Murzin, A.G., A protein catalytic framework with an N-terminal nucleophile is capable of self activation. Nature, 1995. 378(6555): p. 416-19.

8.) Kim, J.K., et al., Crystal structures of glutaryl 7-aminocephalosporanic acid acylase: insight into autoproteolytic activation. Biochemistry, 2003. 42(14): p. 4084-93.

9.) Suresh, C.G., Pundle, A.V., SivaRaman, H., Rao, K.N., Brannigan, J.A., McVey, C.E., Verma, C.S., Dauter, Z., Dodson, E.J., Dodson, G.G., Penicillin V acylase crystal structure reveals new Ntn hydrolase family members. Nature structural biology, 1999. 6(5): p. 414-416.

10.) Rossocha, M., Schultz-Heienbrok, R., Moeller von, H., Coleman, J.P. and Saenger, W., Conjugated Bile Acid Hydrolase Is a Tetrameric N-Terminal Thiol Hydrolase with Specific Recognition of Its Cholyl but Not of Its Tauryl Product. Biochemistry, 2005. 44: p. 5739-5748.

11.) Kim, Y., et al., Precursor structure of cephalosporin acylase. Insights into autoproteolytic activation in a new N-terminal hydrolase family. Journal of Biological Chemistry, 2002. 277(4): p. 2823-9.

12.) Kim, J.K., et al., Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries. Proceedings of the National Academy of Sciences of the United States of America, 2006. 103(6): p. 1732-7.

13.) Michalska, K., Hernandez-Santoyo, A., Jaskolkski, M., The mechanism of autocatalytic activation of plant-type L-asparaginases. JBC, 2008. 283(19): p. 13388-97.

14.) Saarela, J., et al., Autoproteolytic activation of human aspartylglucosaminidase. Biochemical Journal, 2004. 378(Pt 2): p. 363-71.

15.) Mahmood, A., FitzGerald, A.J., Marchbank, T., Ntatsaki, E., Murray, D., Ghosh, S. and Playford, R.J, Zinc carnosine, a health food supplement that stabilizes small bowel integrity and stimulates gut repair processes. Gut, 2007. 56: p. 168-175.

16.) Son, D.O., Satsu, H., Kiso, Y., Totsuka, M. and Shimizu, M., Inhibitory effect of carnosine on interleukin-8 production in intestinal epithelial cells through translational regulation. Cytokine, 2008. 42: p. 265-276.

17.) Hipkiss, A.R., Brownson, C. and Carrier, M.J., Carnosine, the anti-ageing, anti-oxidant dipeptide, may react with protein carbonyl groups. Mechanism of Ageing and Development, 2001. 122: p. 1431-1445.

18.) Geueke, B., Namoto, K., Seebach, D., Kohler, HP., A Novel beta-Peptidyl Aminopeptidase (BapA) from strain 3-2W4 Cleaves Peptide Bonds of Synthetic beta-Tri and beta-Dipeptides. J Bacteriology, 2005. 187(17): p. 5910-17.

19.) Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography, 1993. 26: p. 795-800.

20.) McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, C.C., Read, R.J., PHASER. J. Appl. Crystallogr., 2007. 40: p. 658-74.

21.) Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Iorger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Terwillinger, T.C., PHENIX: Building new software for automated crystallographic structure determination. Acta Cryst., 2002. D(58): p. 1948-54.

22.) Heck, T., Reimer, A., Seebach, D., Gardiner, J., Deniau, G., Lukaszuk, A., Kohler, H-P.E., Geueke, B., β-Aminopeptidase-catalyzed biotransformations of β^2-dipeptides: kinetic resolution and enzymatic coupling. ChemBioChem, 2010. (11): p. 1129-1136

* * * * *

The following two manuscripts describe in detail the structure of the unprocessed (S250A) form of BapA $(\alpha)_4$, the native form $(\alpha\beta)_4$, the residues that are essential for the self-processing, catalytic mechanism and substrate specificity, and a detailed overview of the proposed mechanisms for self-processing and catalytic substrate cleavage is provided. We also solved the structures of ampicillin, enzymatically (β-lactamase) hydrolyzed ampicillin and pefabloc SC (serine peptidase inhibitor) in complex with BapA and suggest the prerequisites for beta-peptidic substrates to be cleaved. In fact, we try to answer the question why BapA specifically cleaves beta-peptides and how this is (mechanistically) achieved, and why BapA does not cleave alpha-peptides. Moreover, the results are compared with data of the previously described DmpA protein, the closest known homologue of BapA.

The manuscripts are formatted as they were submitted to "STRUCTURE".

* * * * *

Crystal structures of the β-aminopeptidase BapA and its unprocessed precursor reveal novel catalytic mechanisms for Ntn hydrolases

(Manuscript 1)

Tobias Merz,[1] Tobias Heck,[2] Birgit Geueke,[2] Dieter Seebach,[3] Hans-Peter E. Kohler,[2] and Markus G. Grütter,[§,1]

[1] Biochemistry Institute, University of Zürich

Winterthurerstrasse 190, 8057 Zürich (Switzerland)

[2] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Department of Environmental Microbiology

Überlandstrasse 133, 8600 Dübendorf (Switzerland)

[3] Laboratory of Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich

Hönggerberg HCI, Wolfgang-Pauli-Strasse 10, 8093 Zürich (Switzerland)

§ corresponding author: Prof. Markus G. Grütter, Biochemistry Institute, University of Zürich Winterthurerstrasse 190, 8057 Zürich (Switzerland); gruetter@bioc.uzh.ch; +41 (0)446355580

HIGHLIGHTS

- BapA of the DmpA-like family is not a “ser-only” β-aminopeptidase
- BapA employs an unusual Glu-Ser-Ser-Glu catalytic tetrad for autoproteolysis and catalysis
- The length and conformation of the active site loop determines the substrate specificity

RUNNING TITLE

Structure of the β-aminopeptidase BapA

SUMMARY

The β-aminopeptidase BapA from *Sphingosinicella xenopeptidilytica* belongs to the *N*-terminal nucleophile (Ntn) hydrolases of the DmpA-like family and has the unprecedented property of cleaving *N*-terminal β-amino-acid residues from peptides. We determined the crystal structures of the native $(\alpha\beta)_4$ heterooctamer and of the 153 kD precursor homotetramer at a resolution of 1.45 Å and 1.8 Å, respectively. Mutational analyses of BapA strongly support a mechanism of autoproteolysis and catalysis based on a tetrad involving the residues Glu133, Ser250, Ser288 and Glu290. Structural and functional data provide insight into the discriminating structural features determining the substrate specificity of DmpA-like enzymes. Our results indicate that the “Ser-only” catalytic mechanism proposed for members of the DmpA-like family needs to be reevaluated.

INTRODUCTION

The class of β-aminopeptidases so far comprises five functionally characterized hydrolytic enzymes, *i.e.* three BapA variants from *Sphingosinicella xenopeptidilytica* 3-2W4 (Geueke, 2005) and *S. microcystinivorans* Y2 (Geueke, 2006) and from *Pseudomonas sp.* MCI3434 (Komeda, 2005) as well as BapF from *Pseudomonas aeruginosa* PAO1 (Fuchs, 2011) and DmpA from *Ochrobactrum anthropi* (Fanuel, 1999b). These enzymes share the exceptional ability to cleave synthetic β-peptides, which consist of backbone-elongated β-amino-acid residues that are not processed by common proteolytic enzymes (Frackenpohl, 2001), (Geueke, 2007), (Wiegand, 2002). Due to their structural properties and high proteolysis resistance, β-peptides are considered promising building blocks for the design of novel peptidomimetics (Aguilar, 2007), (Seebach, 2008). The first and so far only structure of an enzyme with β-aminopeptidase activity is DmpA (Bompard-Gilles, 2000). DmpA has an αββα-sandwich architecture, which is strongly reminiscent of the fold of classical Ntn hydrolases (Brannigan, 1995). However, DmpA is classified as a novel member of the Ntn hydrolase family due to a different connectivity of the secondary structure elements (Bompard-Gilles, 2000), (Oinonen, 2000).

Ntn hydrolases, in general, are synthesized as inactive precursor polypeptides with or without a periplasmic leader sequence. They undergo post-translational autoproteolytic cleavage into heterodimers. The primary nucleophile that initiates this cleavage is either a serine, threonine or cysteine residue, which, upon self-processing, becomes the essential *N*-terminal nucleophile for the catalytic activity. This autoproteolytic activation step is a characteristic trait of all Ntn hydrolases (Brannigan, 1995), (Kim, 2003), (Saarela, 2004), (Michalska, 2008). Various biochemical studies showed that Ntn hydrolases occur in different oligomeric states, such as heterodimers (e.g. penicillin G acylase(McVey, 2001),

heterotetramers (e.g. glutamyltranspeptidase from *Helicobacter pylori* (Boanca, 2006)), heterooctamers (e.g. penicillin V acylase from *Bacillus sphaericus* (Suresh, 1999)) or, exceptionally, as very large assembly in the proteasome (Baumeister, 1997). The substrate specificities of Ntn hydrolases vary considerably. Generally, these enzymes catalyze the hydrolysis of amide bonds. However, the ornithine acetyltransferase (*orf6*) from the clavulanic-acid-biosynthesis gene cluster is an exception. The enzyme has the same fold as DmpA, but catalyzes the reversible transfer of an acetyl group from *N*-acetylornithine to glutamate (Elkins, 2005). Among the Ntn hydrolases, penicillin acylases and β-aminopeptidases have particularly high biocatalytic potential (Bruggink, 1998), (Barends, 2004), (Heck, 2007), (Heck, 2009), (Heck, 2010).

The backbone length of the N-terminal amino acid of the substrate molecule has a strong impact on the enzymatic activity of β-aminopeptidases. Whereas DmpA hydrolyzes D- and L-configured α-amino-acid amides, esters and peptides as well as peptides carrying small *N*-terminal β-amino acids such as β-homoglycine and β^3-homoalanine (Fanuel, 1999a), (Heck, 2006), BapA from *S. xenopeptidilytica* exhibits hydrolytic activity with strict preference for variable *N*-terminal β-amino acids and, remarkably, does not hydrolyze α-peptides (Geueke, 2006), (Heck, 2006). Until now, the natural substrates of β-aminopeptidases remain unknown and structural data about exclusively β-peptide-cleaving enzymes are not available.

Here and in Heck et al. (this issue) we present the first high-resolution crystal structures of the β-aminopeptidase BapA, which has exclusive catalytic activity for β-peptide substrates. Our comprehensive structural and functional study about this novel aminopeptidase includes the crystal structures of the active form, its precursor and of the active form in complex with the serine-peptidase inhibitor AEBSF, ampicillin and ampicillin-

derived penicilloic acid. The analysis of these BapA structures provides insight into the mechanisms of autoproteolysis and substrate processing, and shows the determinants of the enzyme`s substrate specificity and enantiomer selectivity. Moreover, our mutational analysis indicates that BapA, DmpA and other so far not characterized DmpA-like aminopeptidases, do not, as reported, act with a “Ser-only” active-site-residue configuration. Instead of a single catalytically active amino-acid residue, these enzymes seem to employ highly conserved tetrads for the self-processing and the catalytic substrate hydrolysis.

RESULTS

Crystal structure of native BapA

BapA crystallized in the space group $P2_1$ and contained one $(\alpha\beta)_4$ heterooctameric molecule (153 kD) in the asymmetric unit. The high-resolution crystal structure (1.45 Å) of BapA was determined by molecular replacement as described in the Methods section. Each heterodimer comprised a large α-polypeptide chain (residues 1-249) and a small β-polypeptide chain (residues 250-373). The molecule assumed a globular disk-like shape in its precursor form with subunits arranged in a D_2 or 222-symmetry mode, creating three perpendicular two-fold symmetry axes (**Figure 1A**). Each αβ-subunit adopted an αββα-sandwich fold that was formed by two layers of four and five parallel/antiparallel β-sheets flanked by two helices on each side (**Figure 1B**). The connectivity of the secondary structure elements in BapA differed from the classical Ntn hydrolase fold, but it was identical to the fold of DmpA. The total interface area between all subunits comprised 7880.2 $Å^2$ in BapA and suggested a very stable and permanent octameric state in solution that was elucidated experimentally by gel filtration (data not shown). The model was refined to final R/R_{free} values of 13.5% and 15.3%. Crystallographic data are listed in **Table 1.** The BapA molecule contained four active sites, each of which was located at the interface of three adjacent subunits forming the substrate-binding pockets. In analogy to other Ntn hydrolases, Ser250 became the *N*-terminal catalytic nucleophile of the β-polypeptide chain upon precursor processing (**Figure 2A**). Glu133, Ser250, Ser288 and Glu288, that we propose to participate in the catalytic mechanism, were located on the same subunit (**Figure 1A**). A water molecule (B_{av} = 18.1 ±1.1) was in close contact to the carboxyl group of Glu133 in each of the active sites. The shape of the active sites of BapA with their buried catalytic residues excluded the action of associated peptidases that could be involved in the autoproteolytic cleavage.

Crystal structure of the unprocessed BapA precursor (proBapA)

The homotetrameric unprocessed precursor form of BapA (proBapA) was obtained by the replacement of Ser250 for alanine, which prevented the primary cleavage of the enzyme-subunits into α- and β-polypeptides. This mutation allowed the structural investigation of the precursor form, because the native BapA was immediately processed after translation. The enzymatically inactive proBapA crystallized in the space group $P2_12_12_1$ and contained one homotetrameric molecule in the asymmetric unit. The high-resolution crystal structure (1.8 Å) of proBapA was determined by molecular replacement as described in the Methods section. The model was refined to final R/R_{free} values of 15.5% and 18.0%. Crystallographic data are listed in **Table 1**. Overall, the globular shape and the contact area between the subunits strongly resembled BapA. The displacement of the linker connecting the designated α- and β-chains (residues 231 – 249) represented the major structural difference between native BapA and proBapA (**Figure 1C**). In proBapA, this segment, which contained two prolines with dihedral angles in allowed *Ramachandran* regions (Pro240 phi -64.5°/psi -30.9° and Pro245 phi 62.6°/psi 140.6°) either assumed a helical conformation due to partial crystal contacts, or it was largely unstructured. However, in BapA, the same segment assumed a β-sheet-like conformation and the prolines were located in the turns. No further significant deviations from ideal geometry in this linker segment were found. Together with the linker displacement, the re-positioning of Ser250 represented a minor but functionally highly important structural change at the active site. Comparing the distances of the Ala250-backbone-amide nitrogen of proBapA and the α-amino group of Ser250 of native BapA to the OH-group of Ser288, Ser250 rotated around its C-Cα bond upon self-processing causing a reduction of the distance from 3.67 ±0.04 Å (NH-Ala250 – OH-Ser288) to 2.79 ±0.01 Å (NH_3^+-Ser250 – OH-Ser288)

(**Figures 2A and 2B**). The inter-residue distances between Glu133, Glu290 and Ser288 did not significantly change between BapA and proBapA.

Catalytic residues for self-processing and catalysis

We exchanged six amino-acid residues in the vicinity of the catalytic nucleophile Ser250 for alanine residues. The consequences of these mutations on the autoproteolysis of the precursor polypeptide were judged based on the time-dependent appearance of individual α- and β-chains on SDS-PAGE gels. Native BapA was already fully processed after purification, while the BapA mutants did not self-process at all or only partially within 192 hours (**Table 2**). In fact, the exchange of Ser250, Ser288 or Glu290 for alanine completely abrogated precursor cleavage: neither the individual α- and β-polypeptide chains could be detected on SDS-PAGE gels (**Figure 3**) nor did the mutant enzymes exhibit any residual enzymatic activity after 192 hours of incubation (**Table 2**). The BapA mutants E133A, K248A and N249A were clearly delayed in self-processing (SDS-PAGE for K248A not shown). Nevertheless, activity measurements after 192 hours of incubation revealed that K248A and N249A mutants were as active as native BapA, but the exchange of Glu133 completely abolished any activity (**Table 2**).

Because the BapA mutants E133A, S250A, S288A and E290A did not undergo the complete processing step, the effects of their mutations on substrate binding and conversion could not be investigated. In order to study the mechanism of substrate cleavage independently from precursor processing, we co-expressed the individual α- and β-polypeptide chains of BapA from a Duet expression vector and introduced the same mutations that were previously investigated for their effects on self-processing. We ensured by sequence analysis that the

newly introduced *N*-terminal methionine residue of the β-polypeptide chain was posttranslationally removed by *E.coli*. The cell-free extracts containing the respective BapA Duet mutants were then tested for their catalytic activity with the substrate H-β^3hAla-*p*NA. The mutants E133A Duet and S250A Duet were inactive, whereas the mutants S288A Duet and E290A Duet exhibited <1% and 24% activity, respectively, compared to wild-type BapA Duet (**Table 2**).

Mechanism of self-processing

Based on the structural data and our mutational studies, we propose the following self-processing mechanism: Glu133 represents the only amino acid in the vicinity of Ser250 to enhance the nucleophilicity of Ser250-Oγ. It supposedly acts as a base and initiates the first step of the cleavage process through the catalytic water molecule (**Figure 4A,** precursor state). However, the residual autoprocessing ability of the BapA E133A mutant indicates that the activation of Ser250 can be partially circumvented by water molecules and hydroxyl ions from the environment (**Figure 3**). We speculate that the nucleophilic attack of Ser250-Oγ on the preceding peptide bond and the subsequent formation of the oxy-oxazolidine intermediate is facilitated by the ε-amino group of Lys248, which points towards the active-site pocket and reversibly acts as proton donor to the carbonyl group of Asn249 (**Figure 2B**). This reversible protonation might influence the rate of the autoproteolytic cleavage by promoting the formation of the oxy-oxazolidine intermediate state. Moreover, the scissile peptide bond is stabilized by a salt bridge formed between Glu133 and Lys248. Transitions within the oxy-oxazolidine intermediate state, which are characterized by a change from a peptide bond to an ester bond, lead to the formation of the acyl enzyme (**Figure 4A**, steps 1 and 2). Finally, the

N-terminal nucleophile Ser250 is exposed through reversible protonation steps within tetrahedral intermediates that are stabilized by the ε-amino group of Lys248 and the amide proton of Gly289 (**Figure 4A**, steps 3 and 4). Overall, proBapA requires the four residues Glu133, Ser250, Ser288 and Glu290 in order to be fully processed whereas Lys248 and Asn249 fulfill accessory functions.

Catalytic substrate-cleavage mechanism

The substrate molecule freely diffuses into the active-site pocket of BapA and a salt bridge is formed between its amino terminus and Glu133 of BapA (**Figure 4B**, Michaelis complex). Glu290 supposedly activates Ser250-Oγ for the nucleophilic attack on the carbonyl group of the substrate, while Ser288-OH and the α-amino group of Ser250 enable the proton transport of this event chain. Leu135-NH and Asn207-NH_2 stabilize the negative charge of the oxyanion of the tetrahedral intermediate (**Figure 4B**, step 1), which collapses to the acyl enzyme under release of the *C*-terminal substrate moiety. Glu133 subsequently activates the catalytic water molecule, which hydrolyzes the acyl enzyme (**Figure 4B**, step 2). The amino-acid configuration of the BapA active site generally offers two pathways for the activation of Ser250-Oγ and the subsequent hydrolysis of the acyl enzyme. Instead of the previously described Ser250-Ser288-Glu290 triad, Ser250-Oγ could alternatively be activated by Glu133 and the catalytic water molecule for the initial nucleophilic attack on the carbonyl group of the substrate. In this case, the hydrolysis of the acyl enzyme would subsequently be initialized by Glu290 (**Figure S1**). Similarly to the last steps during self-processing, reversible protonation steps within tetrahedral intermediates lead to the rupture of the acyl enzyme and to the release of the substrate`s *N*-terminal amino acid (**Figure 4B**, steps 3 and 4). Overall, the self-

processing and the catalytic cleavage mechanisms are very similar. However, they differ between the protonation states of Glu133. During self-processing, Glu133 is deprotonated upon formation of the acyl enzyme and remains deprotonated until the release of the linker segment. On the other hand, the protonation of Glu133 during substrate hydrolysis breaks the salt-bridge interaction with the hydrolyzed, but still covalently bound *N*-terminal substrate moiety and enables its unimpaired release upon hydrolysis of the acyl enzyme. In summary, BapA hydrolyzes substrates with a catalytic tetrad composed of the residues Glu133, Ser250, Ser288 and Glu290. Noteworthy, Ser250 involves its OH-group and its α-amino group during the catalytic cycles.

Comparison between the substrate-binding pockets of BapA and DmpA

The superposition of the structures of native BapA, DmpA (PDB ID 1b65) and a so far uncharacterized protein from *Pyrococcus horikoshii* (PH0078, PDB ID 2drh) showed a good overlap of the corresponding catalytic residues. However, the sizes of the substrate-binding pockets differed notably (**Figure 5**). In BapA, a 3_{10}-helix comprising the residues Glu120 to Arg126 created a wide substrate-binding pocket, which was largely occluded by loops in DmpA (Gln131 to Trp137) and PH0078 (Glu120 to Ser128). From this superposition we concluded that Trp137, which pointed into the active-site cavity of DmpA, restricted catalysis to the removal of sterically undemanding *N*-terminal amino acids. To confirm this hypothesis, we exchanged Trp137 for alanine by site-directed mutagenesis in order to widen the substrate-binding pocket of DmpA and thus make it accessible for bulkier substrates. Hereby, it is important to mention that the point mutation W137A negatively affected the self-processing rate of the DmpA precursor polypeptide into the individual α- and β-polypeptides. In contrast

to the native BapA and DmpA proteins that were fully processed after the purification procedure, DmpA W137A was only half-processed after purification and three days of incubation at pH 7.2 and 37°C. A survey of the kinetic parameters of the conversion of the β^3-homoamino-acid *p*-nitroanilides H-βhGly-*p*NA, H-β^3hAla-*p*NA, H-D-β^3hAla-*p*NA, H-β^3hPhe-*p*NA and H-D-β^3hPhe-*p*NA by BapA, DmpA and DmpA W137A is shown in **Table 3**. While BapA converted the aromatic β^3-homophenylalanine *p*-nitroanilides of L- and D-configuration with k_{cat}/K_M values of 14,500 and 940 $M^{-1}s^{-1}$, respectively, the conversion of these two enantiomers by DmpA was negligible (k_{cat}/K_M < 10 $M^{-1}s^{-1}$). The single amino-acid exchange of Trp137 for alanine in DmpA drastically altered the substrate specificity of DmpA W137A leading to more than 280- and 670-fold increased catalytic efficiencies for the conversion of the substrates H-β^3hPhe-*p*NA and H-D-β^3hPhe-*p*NA, respectively. On the other hand, DmpA W137A converted substrates with small side chains, *i.e.* H-βhGly-*p*NA, H-β^3hAla-*p*NA and H-D-β^3hAla-*p*NA, less efficiently than wild-type DmpA and displayed overall k_{cat}/K_M-values similar to BapA for these substrates (**Table 3**).

Sequence comparison of putative β-aminopeptidases and implications on their substrate specificity

In addition to the five functionally characterized β-aminopeptidases (Geueke, 2006), (Komeda, 2005), (Fuchs, 2011), (Fanuel, 1999a), (Heck, 2006), a BLAST search revealed a large number of protein sequences showing high similarity to the sequence of BapA from *S. xenopeptidilytica*. They mainly originated from Gram-negative bacteria, among them many pathogenic species such as *Burkholderia mallei, B. pseudomallei* and various strains of *Pseudomonas aeruginosa*, but also from archaea (e.g *Pyrococcus horikoshii*), fungi (e.g

Aspergillus oryzae) and yeast (e.g *Yarrowia lipolytica*). Strikingly, the region downstream of the *N*-terminal catalytic nucleophile (ranging from Ser250 to Ser290) and the region around the salt-bridge-forming residue Glu133 were highly conserved among all compared sequences, indicating that these regions are crucial for catalysis (**Figure 6**). Glutamate at position 290 of BapA was less, but functionally conserved. In contrast to these highly conserved regions, the compared sequences showed great variability over a stretch of 15 to 20 amino acids upstream of Glu133. As shown in **Figure 5**, this sequence stretch corresponded to the active-site loop (3_{10}-helix in BapA), which defined the width of the enzymes` substrate-binding pockets and their substrate specificities (**Table 3**).

DISCUSSION

In this study, we report the crystal structures of the native and precursor form of BapA from *S. xenopeptidilytica*. Native BapA revealed a heterooctameric organization comprising four heterodimers in a globular D_2-symmetry arrangement. Each heterodimer assumed an αββα-sandwich fold, which was very similar to the core-fold of classical Ntn hydrolases (Brannigan, 1995), (Oinonen, 2000). ProBapA revealed a homotetrameric organization with the same globular arrangement and core-fold. The displacement of the linker segment between the α- and β-subunits (residues 231 – 249) represented the major structural difference between native BapA and proBapA. Although this segment contained two prolines in allowed *Ramachandran* regions and no further significant deviations from ideal geometry, we expect these prolines to provide the initial driving force for the large conformational change from a helical to a β-sheet-like conformation in the autoprocessing step. Similar observations of strained conformations of linker sequences with prolines that are supposed to regulate the autoprocessing of Ntn hydrolases are reported for the Ntn hydrolases glutaryl 7-aminocephalosporanic acid acylase (Kim, 2003) and cephalosporin acylase (Kim, 2006).

BapA differed from other classical Ntn hydrolases by the connectivity of the secondary structure elements, but had the same fold as DmpA from *O. anthropi* (Bompard-Gilles, 2000). Therefore, BapA belongs to the DmpA-like family. DmpA is reported to be closely related to the superfamily of Ntn hydrolases and to exhibit a single-residue catalytic mechanism, in which the *N*-terminal nucleophile (Ser250), incorporated in a β-sheet, uses its own α-amino group as Brønsted base for activation (Bompard-Gilles, 2000).

BapA and DmpA shared 41.8% sequence identity and were structurally and biochemically closely related. The catalytic residues of BapA overlapped extremely well with corresponding conserved amino acids of DmpA, although their substrate specificities vary

(Heck, 2006). We therefore propose that substrate conversions by DmpA and BapA involve a similar mechanism. We showed that the residues Glu133, Ser250, Ser288 and Glu290 of BapA (corresponding to Glu144, Ser250, Ser288 and Asp290 of DmpA) have important functions for the self-processing and the catalytic mechanism. The crystal structure of native BapA revealed short hydrogen-bond distances of 2.8 Å between Ser250-NH_3^+ and Ser288-OH and of 2.6 Å between Ser288-OH and the carboxyl group of Glu290, suggesting that the catalytic nucleophile Ser250-Oγ is activated by Glu290 and Ser288. These assumptions are in accordance with our kinetic analysis that showed 100-fold and 4-fold decreased activities, when Ser288 and Glu290 were exchanged for alanines, respectively. Nevertheless, the residual catalytic activity of the S288A Duet mutant indicates that either Ser250 is indeed capable of self-activation as postulated by (Bompard-Gilles, 2000), or that Glu133 exhibits very weak activity. In comparison the native enzyme, Glu133 and Ser250 of the S288A Duet mutant only seem to play a minor role for the activation of the catalytic nucleophile. In fact, this loss of activity is not surprising. We assume the α-amino group of Ser250 to be mostly protonated (pK_a > 9). Therefore, the amino group is unlikely to act as a base abstracting a proton from Ser250-OH and to provide the major driving force for the activation of its own nucleophile.

The major consequence arising with our study is that the proposed “Ser-only” catalytic mechanism of DmpA must be reconsidered. Our results strongly suggest that BapA and DmpA feature catalytic tetrads (Glu-Ser-Ser-Glu and Glu-Ser-Ser-Asp, respectively) rather than single catalytic serine residues. This clearly extends the currently proposed catalytic mechanism of DmpA. The importance of these catalytic residues for BapA, DmpA and other DmpA-like proteins is further supported by their high conservation on the structural (**Figure 5**) and on the sequence level (**Figure 6**). Interestingly, despite the close structural similarity of

the αββα-sandwich fold, the Ntn hydrolase superfamily comprises enzymes with different catalytic mechanisms. The prototypic Ntn hydrolase penicillin G acylase is reported to have a single catalytic serine residue (Duggleby, 1995), the γ-glutamyltranspeptidase from *H. pylori* reveals a threonine-threonine catalytic dyad (Boanca, 2007) and the 20S proteasome from *T. acidophilum* employs a catalytic tetrad with a threonine residue as primary nucleophile (Löwe, 1995).

The superposition of the available crystal structures of β-aminopeptidases and a comparison of the biochemically characterized enzymes with related amino-acid sequences showed that despite the highly conserved positions of the catalytic amino-acid residues, the size and chemical environment of the substrate-binding pocket largely differs among these proteins. We expect a non-conserved sequence stretch, assuming different conformations within the substrate-binding pocket, to mainly determine the substrate specificity of proteins of the DmpA-like family. The wide substrate-binding pocket of BapA allows the enzyme to catalyze reactions of peptides with bulky *N*-terminal β-amino-acid residues, while Trp137 in a loop occludes this pocket and restricts DmpA to hydrolyze substrates with sterically undemanding *N*-terminal amino acids. However, by exchanging Trp137 for alanine, DmpA could be rationally engineered to an altered enzyme with catalytic properties that are more similar to BapA than to DmpA. Another interesting candidate to study the influence of modifications within the substrate-binding pocket on catalysis is the biochemically non-characterized protein PH0078 from *P. horikoshii* (PDB ID 2drh). The active-site loop of PH0078 assumes a similar conformation within the substrate-binding pocket like the corresponding loop of DmpA but exposes a sterically less demanding serine residue instead of a tryptophane towards the active-site residues. Therefore, we expect PH0078 to exhibit a substrate profile, which is similar to the one of the DmpA mutant W137A. Further specific

amino-acid exchanges in this particularly variable region might lead to the creation of β-aminopeptidases with interesting novel specificities. The structure-based rational design of modified β-aminopeptidases represents a promising approach to improve the properties of the available enzymes with respect to biocatalytic applications, such as the enzyme-catalyzed synthesis of β-peptides or the production of enantiopure β-amino acids.

EXPERIMENTAL PROCEDURES

Generation of the BapA Duet expression constructs and mutants

The gene sequences coding for the individual α- and β-polypeptide chains of BapA were amplified by PCR from the plasmid p3BapA containing *bapA* from *S. xenopeptidilytica* 3-2W4 without its periplasmic signal sequence (Geueke, 2006) and separately cloned into a pETDuet-1 vector (Novagen, Madison, USA). Sequences of primers and information about site-directed mutagenesis can be found in the **supplemental data**.

Expression and purification of BapA for biochemical assays and protein crystallization

The production and purification of the β-aminopeptidase BapA from *S. xenopeptidilytica* 3-2W4 and DmpA from *O. anthropi* was done following established procedures (Geueke, 2005), (Geueke, 2006), (Fanuel, 1999b). All mutants of BapA and DmpA were purified accordingly to purity levels > 70%. Cell-free extracts and partially purified proteins were analyzed by SDS-PAGE using pre-cast 10% Novex tricine gels (Invitrogen AG, Basel, Switzerland) according to the manufacturer`s instructions. The protein gels were stained with Coomassie Brilliant Blue and the relative intensities of the protein bands were determined with a GS-800 calibrated imaging densitometer and the software Quantity One (Bio-Rad, Reinach, Switzerland). The β-polypeptide chain of the partially purified BapA S250 Duet was transferred from the gel to PVDF membranes by electroblotting and the N-terminus was sequenced at the Functional Genomics Center Zürich (Switzerland).

Before crystallization, the purified protein was dialyzed against a 0.5 mM Tris/HCl buffer (pH 8) and concentrated to 15 mg/ml prior to crystallization. Crystals were grown in 2 µl drops with a 1:1 ratio of mother liquor (1.8 M ammonium sulfate and 100 mM Tris/HCl, pH 8.0) to

purified BapA in 24-well Cryschem sitting-drop plates (Hampton Research, Aliso Viejo, CA, USA) using the vapor-diffusion method at 20°C. Within one day, crystals of BapA grew as stacked layers of thin plates. Sequential macroseedings with crystal fragments that were transferred to new wells containing 1.5 M ammonium sulfate, 100 mM Tris/HCl (pH 8.0), yielded single three-dimensional, mostly plate-shaped crystals of up to 300 × 300 × 30 μm in size. Cryo-protected crystals (mother liquor and 30% glycerol) were flash-frozen in liquid nitrogen.

Structure determination

Data were collected at the Swiss Light Source beamline PX on a 6M Pilatus detector in a cryostream at 100 K at 1 Å wavelength. The data sets were indexed, integrated and scaled with XDS (Kabsch, 1993). The structure of BapA was solved by molecular replacement with PHASER (McCoy, 2007) using one subunit (α- and β-chain) of the heterooctameric structure of DmpA (PDB ID 1b65). The search probe was additionally tailored at the *C*-terminus of the α-chain (residues 224 to 245). The structure of proBapA was solved by molecular replacement with PHASER using one subunit (α- and β-chain) of native BapA. The *C*-terminal tail of the α-chain (residues 231-249) was omitted in this search probe. Unambiguous solutions of proBapA and native BapA were obtained, including an extended region of positive difference electron density (Fo-Fc) in the initial map (model phases of BapA and DmpA, respectively) that corresponded to the omitted amino acid linker sequences. Each polypeptide chain comprised 373 amino acids. The final model of proBapA contained residues 1-236 / 248-371 for chain A, 1-236 / 248-317 for chain B, 1-238 / 248-371 for chain C and 1-371 for chain D. The final model of native BapA contained residues 1-245 / 250-371

for chain A, 1-243 / 250-371 for chain B, 1-245 / 250-371 for chain C and 1-239 / 250-371 for chain D. The refinement of the structures was carried out through multiple cycles of manual adjustments and rebuilding using COOT (Emsley, 2004), and refinement was carried out using PHENIX (Adams, 2010). Interface area calculations were performed using the Pisa server v.1.18. Superpositions for figures were done using SSM (Krissinel, 2004). Figures were created with the program Pymol (DeLano).

Activity assays

The β-aminopeptidase-catalyzed conversion of the β^3-homoamino-acid *p*-nitroanilides H-βhGly-*p*NA, H-β^3hAla-*p*NA, H-D-β^3hAla-*p*NA, H-β^3hPhe-*p*NA and H-D-β^3hPhe-*p*NA was determined spectrophotometrically by monitoring the release of free *p*-nitroaniline at a wavelength of 405 nm (ε_{405} = 8,800 $M^{-1}cm^{-1}$) for one minute. For the determination of kinetic constants, the assay mixtures contained different molar concentrations of the substrates in a 100 mM potassium phosphate buffer at pH 7.2 at 37°C; 10% DMSO was added as co-solvent. The standard assay contained 5 mM H-β^3hAla-*p*NA as substrate. The reactions were initiated by the addition of one of the purified enzymes BapA, DmpA or DmpA W137A, or by the addition of cell-free extract containing recombinantly expressed enzyme. The initial rates of the enzyme-catalyzed release of *p*-nitroanilide were calculated and fitted to the Michaelis-Menten model by non-linear regression analysis with the VisualEnzymics software (Softzymics, Princeton, NJ, USA) for the program IGOR Pro (WaveMetrics, Oswego, OR, USA).

PDB ACCESSION NUMBERS

The structures of proBapA and native BapA and proBapA are accessible under the PDB ID 3n2w and 3n5i, respectively.

ACKNOWLEDGEMENTS

We thank the staff of beamline PX of the Swiss Synchrotron Light Source (PSI) in Villigen, Switzerland for excellent technical assistance. This project was funded by the Swiss National Science Foundation (grants to M.G.G.) and the Swiss NCCR Structural Biology program. T.H. was supported by the Deutsche Bundesstiftung Umwelt (DBU Project No. 13176-32), B.G was supported by the Swiss National Science Foundation (SNF Project No. 3152A0-100770).

AUTHOR CONTRIBUTIONS

H.P.K. and B.G. conceived the work. T.M. carried out crystallographic data collection, structure determination and refinement. T.H. and B.G. carried out protein expression, purification and crystallization experiments. T.H. carried out mutations and kinetic analysis. D.S. contributed to the enzyme mechanism and provided β-peptidic substrates. T.M. and T.H. wrote the paper and prepared the figures. D.S., M.G.G., B.G. and H.P.K. contributed to discussion, data interpretation and manuscript preparation. M.G.G., B.G. and H.P.K. supervised the work.

COMPETING FINANCIAL INTERESTS

We declare no competing financial interests.

REFERENCES

Adams, P.D. et al., (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D: Biol. Crystallogr. *66*, 213-221.

Aguilar, M.-I., Purcell, A.W., Devi, R., Lew, R., Rossjohn, J., Smith, A.I., and Perlmutter, P. (2007). β-Amino acid-containing hybrid peptides - new opportunities in peptidomimetics. Org. Biomol. Chem. *5*, 2884-2890.

Barends, T.R.M., Yoshida, H., and Dijkstra, B.W. (2004). Three-dimensional structures of enzymes useful for β-lactam antibiotic production. Curr. Opin. Struct. Biol. *15*, 356-363.

Baumeister, W., and Lupas, A. (1997). The proteasome. Curr. Opin. Struct. Biol. *7*, 273-278.

Boanca, G., Sand, A., and Barycki, J.J. (2006). Uncoupling the enzymatic and autoprocessing activities of *Helicobacter pylori* γ-glutamyltranspeptidase. J. Biol. Chem. *281*, 19029-19037.

Boanca, G., Sand, A., Okada, T., Suzuki, H., Kumagai, H., Fukuyama, K., and Barycki, J.J. (2007). Autoprocessing of *Helicobacter pylori* γ-glutamyltranspeptidase leads to the formation of a threonine-threonine catalytic dyad. J. Biol. Chem. *282*, 534-541.

Bompard-Gilles, C., Villeret, V., Davies, G.J., Fanuel, L., Joris, B., Frère, J-M., and Beeumen van, J. (2000). A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-Ala-esterase/ amidase from *Ochrobactrum anthropi*. Structure *8*, 153-162.

Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C.E., Smith, J., Tomchick, D.R., and Murzin, A.G. (1995). A protein catalytic framework with an *N*-terminal nucleophile is capable of self-activation. Nature *378*, 416-419.

Bruggink, A., Roos, E.C., and Vroom de, E. (1998). Penicillin acylase in the industrial production of β-lactam antibiotics. Org. Process Res. Dev. *2*, 128-133.

DeLano, W.L. The pymol molecular graphics system. DeLano Scientific: San Carlos, CA, USA, 2002.

Duggleby, H.J., Tolley, S.P., Hill, C.P., Dodson, E.J., Dodson, G., and Moody, P.C.E. (1995). Penicillin acylase has a single-amino-acid catalytic centre. Nature *373*, 264-268.

Elkins, J.M., Kershaw, N.J., and Schofield, C.J. (2005). X-ray crystal structure of ornithine acetyltransferase from the clavulanic acid biosynthesis gene cluster. Biochem. J. *385*, 565-573.

Emsley, P., and Cowtan, K. (2004). COOT: model-building tools for molecular graphics. Acta Crystallogr. Sect. D: Biol. Crystallogr. *60*, 2126-2132.

Fanuel, L., Goffin, C., Cheggour, A., Devreese, B., Driessche van, G., Joris, B., Beeumen van, J., and Frère, J-M. (1999a). The DmpA aminopeptidase from *Ochrobactrum anthropi*

LMG7991 is the prototype of a new terminal nucleophile hydrolase family. Biochem. J. *341*, 147-155.
Fanuel, L., Thamm, I., Kostanjevecki, V., Samyn, B., Joris, B., Goffin, C., Brannigan, J., Beeumen Van, J., and Frère, J.M. (1999b). Two new aminopeptidases from *Ochrobactrum anthropi* active on D-alanyl-p-nitroanilide. Cell. Mol. Life Sci. *55*, 812-818.

Frackenpohl, J., Arvidsson, P.I., Schreiber, J.V., and Seebach, D. (2001). The outstanding biological stability of β- and γ-peptides toward proteolytic enzymes: an *in vitro* investigation with fifteen peptidases. ChemBioChem *2*, 445-455.

Fuchs, V., Jaeger, K-E., Wilhelm, S., and Rosenau, F. (2011). The BapF protein from *Pseudomonas aeruginosa* is a β-peptidyl aminopeptidase. World J. Microbiol. Biotechnol. *27*, 713-718.

Geueke, B., Heck, T., Limbach, M., Nesatyy, V., Seebach, D., and Kohler, H.-P.E. (2006). Bacterial β-peptidyl aminopeptidases with unique substrate specificities for β-oligopeptides and mixed β,α-oligopeptides. FEBS J. *273*, 5261-5272.

Geueke, B., and Kohler, H.-P.E. (2007). Bacterial β-peptidyl aminopeptidases: on the hydrolytic degradation of β-peptides. Appl. Microbiol. Biotechnol. *74*, 1197-1204.

Geueke, B., Namoto, K., Seebach, D., and Kohler, H.-P.E. (2005). A novel β-peptidyl aminopeptidase (BapA) from strain 3-2W4 cleaves peptide bonds of synthetic β-tri- and β-dipeptides. J. Bacteriol. *187*, 5910-5917.

Heck, T., Kohler, H.-P.E., Limbach, M., Flögel, O., Seebach, D., and Geueke, B. (2007). Enzyme-catalyzed formation of β-peptides: β-peptidyl aminopeptidases BapA and DmpA acting as β-peptide-synthesizing enzymes. Chem. Biodiversity *4*, 2016-2030.

Heck, T., Limbach, M., Geueke, B., Zacharias, M., Gardiner, J., Kohler, H.-P.E., and Seebach, D. (2006). Enzymatic degradation of β- and mixed α,β-oligopeptides. Chem. Biodiversity *3*, 1325-1348.

Heck, T., Reimer, A., Seebach, D., Gardiner, J., Deniau, G., Lukaszuk, A., Kohler, H.-P.E., and Geueke, B. (2010). β-aminopeptidase-catalyzed biotransformations of β^2-dipeptides: kinetic resolution and enzymatic coupling. ChemBioChem *11*, 1129-1136.

Heck, T., Seebach, D., Osswald, S., Wiel ter, M.K.J., Kohler, H.-P.E., and Geueke, B. (2009). Kinetic resolution of aliphatic β-amino acid amides by β-aminopeptidases. ChemBioChem *10*, 1558-1561.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795-800.

Kim, J.K., Yang, I.S., Rhee, S., Dauter, Z., Lee, Y.S., Park, S.S., and Kim, K.H. (2003). Crystal structures of glutaryl 7-aminocephalosporanic acid acylase: insight into autoproteolytic activation. Biochemistry *42*, 4084-4093.

Kim, J.K., Yang, I.S., Shin, H.J., Cho, K.J., Ryu, E.K., Kim, S.H., Park, S.S., and Kim, K.H. (2006). Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries. Proc. Natl. Acad. Sci. U. S. A. *103*, 1732-1737.

Komeda, H., and Asano, Y. (2005). A DmpA-homologous protein from *Pseudomonas sp.* is a dipeptidase specific for β-alanyl dipeptides. FEBS J. *272*, 3075-3084.

Krissinel, E., and Henrick, K. (2004). Secondary structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. Sect. D: Biol. Crystallogr. *60*, 2256-2268.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. Science *268*, 533-539.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). PHASER crystallographic software. J. Appl. Crystallogr. *40*, 658-674.

McVey, C.E., Walsh, M.A., Dodson, G.G., Wilson, K.S., and Brannigan, J.A. (2001). Crystal structures of Penicillin acylase enzyme-substrate complexes: structural insights into the catalytic mechanism. J. Mol. Biol. *313*, 139-150.

Michalska, K., Hernandez-Santoyo, A. & Jaskolski, M. (2008). The mechanism of autocatalytic activation of plant-type L-asparaginases. J Biol Chem *283*, 13388-13397.

Oinonen, C., and Rouvinen, J. (2000). Structural comparison of Ntn-hydrolases. Protein Sci. *9*, 2329-2337.

Saarela, J., Oinonen, C., Jalanko, A., Rouvinen, J., and Peltonen, L. (2004). Autoproteolytic activation of human aspartylglucosaminidase. Biochem. J. *378*, 363-371.

Seebach, D., and Gardiner, J. (2008). β-peptidic peptidomimetics. Acc. Chem. Res. *41*, 1366-1375.

Suresh, C.G., Pundle, A.V., SivaRaman, H., Rao, K.N., Brannigan, J.A., McVey, C.E., Verma, C.S., Dauter, Z., Dodson, E.J., and Dodson, G.G. (1999). Penicillin V acylase crystal structure reveals new Ntn hydrolase family members. Nature *6*, 414-416.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673-4680.

Wiegand, H. et al., (2002). The outstanding metabolic stability of a ^{14}C-labeled β-nonapeptide in rats - *in vitro* and *in vivo* pharmacokinetic studies. Biopharm. Drug Dispos. *23*, 251-262.

FIGURE LEGENDS

Figure 1: Structural overview of native BapA and proBapA

(A) Cartoon representation of the $(\alpha\beta)_4$-BapA heterooctamer viewed along the P axis (perpendicular to the paper plane). The Q and R axes are shown as dotted lines. The individual subunits A, B, C and D, each comprising an α- and a β-polypeptide chain, are shown in cyan, blue, orange and red, respectively. The blow-up frame shows the catalytic residues Glu133, Ser250, Ser288 and Glu290 of the subunit A as sticks in yellow and the residues of the substrate-binding pocket as sticks in grey (subunit A: Thr76, Thr100, Leu135, Leu287; subunit C: Thr316; subunit D: Leu84, Val88, Gln90, Leu92, Phe124, Ser125, Leu127, Leu128, Leu303).

(B) Cartoon representation of one αβ-subunit. The secondary structure elements contributing to the Ntn fold are colored in red and blue.

(C) Active-site pocket of BapA before and after the autoproteolytic cleavage: The catalytic residues Glu133, Ser250, Ser288 and Glu290 are shown in yellow. The linker segment (residues 231 to 249) is shown in its precursor (magenta) and processed form (blue). The residues Pro240 and Pro245 were located within a helical loop conformation in the precursor, which upon autoproteolytic processing assumed a β-sheet-like conformation. Hydrogen bonds are indicated with dotted lines; the given distances refer to the linear Cα-shifts of Pro240 and Pro245.

Figure 2: Close-up view into the active sites of native BapA and proBapA

(A) Native BapA: rotation of Ser250 around the C-Cα-bond decreased the $NH_3^+{}_{250}$-OH_{288} distance from 3.7 Å to 2.8 Å. The 2fofc electron density map is contoured at 1.5 σ.

(B) BapA precursor mutant S250A (proBapA): The catalytic residues Glu133, Ser288, Glu290 and Ser250/Ala250 are shown as sticks in yellow and the catalytic water molecule is shown as a red sphere. The carboxy-terminal residues of the α-chain before the autoproteolytic cleavage are shown in magenta. Distances (indicated by dotted arrows) of the ε-amino-group of Lys248 to the carboxyl-group of Glu133 and the carbonyl-group of Asn249 vary between the subunits from 3.6 – 5.2 Å and 3.0 – 5.5 Å, respectively.

Figure 3: SDS-PAGE analysis of five processing-deficient or slow-processing BapA point mutants.

The mutant proteins (0.6 mg/ml) were incubated in 50 mM Tris/HCl buffer (pH 8) at 37°C. Samples were taken after 0 h (0), 54 h (1), 120 h (2) and 192 h (3).

Figure 4: Schematic drawings of the self-processing and the catalytic substrate-cleavage mechanisms.

(A) Formation of the oxy-oxazolidine through the activation of Ser250-Oγ by Glu133 and the catalytic water molecule to allow for nucleophilic attack on the preceding peptide bond to Asn249. Step 1: Indirect protonation of the secondary amino group of the oxy-oxazolidine intermediate by Glu133, presumably influenced by Glu290. Stabilization of the oxyanion by the backbone amide group of Gly289 and the ε-amino group of Lys248. Step 2: Collapse of

the tetrahedral intermediate into the acyl enzyme and subsequent attack of the acyl enzyme by the activated catalytic water molecule, initiated by Glu290. Step 3: Indirect protonation of the amino group of Ser250 by Glu290 and Ser288 and rupture of the tetrahedral intermediate through subsequent proton and electron transfer reactions. Step 4: Release of the carboxy-terminal tail of the α-polypeptide chain and protonation of the amino group of Ser250.

(B) Michaelis complex: Formation of a salt bridge between Glu133 and the substrate`s amino terminus. Activation of Ser250-Oγ by Glu290, and initiation of the nucleophilic attack on the carbonyl group of the substrate. Step 1: Stabilization of the oxyanion by the backbone amide group of Leu135 and the δNH_2-group of Asn207. Substrate cleavage and formation of the acyl enzyme through proton transfers within the tetrahedral intermediate. Step 2: Activation of the catalytic water molecule by Glu133 and initiation of the acyl-enzyme hydrolysis. Step 3: Indirect protonation of the amino group of Ser250 by Glu290 and Ser288 and rupture of the tetrahedral intermediate through subsequent proton and electron transfer reactions. The salt-bridge interaction between Glu133 and the substrate`s amino terminus is abolished. Step 4: Release of the cleaved product and reconstitution of the active-site residues for another catalytic cycle.

Solid arrows denote electron-pair shifts, dashed arrows denote catalytic influence; hydrogen bonds are indicated by dotted lines. R_{249} = Asn249, R_{251} = Leu251, R = *N*-terminal amino-acid residue of the substrate, Rx = side chain of the *N*-terminal amino acid, Ry = *C*-terminal substrate moiety.

Figure 5: Close-up view into the active-site pocket of native BapA and superposition of BapA with DmpA and PH0078.

Stereo representation of the active sites of DmpA from *O. anthropi* (PDB ID 1b65) and a non-characterized protein form *P. horikoshii* (PH0078, PDB ID 2drh), superposed onto BapA from *S. xenopeptidilytica* and viewed 180° rotated around the *Q*-axis (*cf*. **Figure 1A**). The active-site loops containing Ser128 (PH0078) and Trp137 (DmpA) are part of the highly variable sequence stretch of 15 to 20 amino acids upstream of the conserved Glu133 (BapA numbering, see **Figure 6**). The subunits A and D of BapA are shown in cyan and red, the active-site loops as well as the residues of DmpA and PH0078 are shown as sticks in orange and violet, respectively. The active-site residues Glu133, Ser250, Ser288 and Glu290 of BapA are shown as sticks in yellow. They correspond to Glu135/Glu144, Ser239/Ser250, Ser277/Ser288 and Asp279/Asp290 of PH0078 and DmpA, respectively.

Figure 6: Partial alignment of β-aminopeptidase-like sequences

Sequence alignment of BapA from *S. xenopeptidilytica* with four previously characterized β-aminopeptidases (2-5) and selected, non-characterized protein sequences that were retrieved from a BLAST search (6-18). Two sections of the alignment are shown (amino acids 104 – 145 and 250 – 290). Identical amino acids are highlighted in black, similar amino acids are highlighted in grey. The numbering refers to BapA from *S. xenopeptidilytica*. Trp137 of DmpA from *O. anthropi*, which was selected as a target for mutagenesis, and Ser128 of the unknown protein PH0078 from *P. horikoshii* (NP_142096) are underlined and printed in bold. The alignment was created with the program ClustalW (Thompson). (1) BapA from *Sphingosinicella xenopeptidilytica* 3-2W4 (AAX93858), (2) DmpA from *Ochrobactrum*

anthropi (CAA66259), (3) BapA from *Sphingosinicella microcystinivorans* ABC59253, (4) BapA from *Pseudomonas sp*. MCI3434 (BAE02664), (5) BapF from *Pseudomonas aeruginosa* PAO1 (NP_250177). Non-characterized protein sequences from (6) *Agrobacterium tumefaciens* C58, (7) *Burkholderia mallei* ATCC 23344, (8) *Mycobacterium gilvum* PYR-GCK, (9) *Mycobacterium smegmatis* MC2 155, (10) *Myxococcus xanthus* DK 1622, (11) *Photorhabdus asymbiotica*, (12) *Pseudomonas entomophila* L48, (13) *Pseudomonas fluorescens* Pf-5, (14) *Pseudomonas putida* KT2440, (15) *Stigmatella aurantiaca* DW4/3-1, (16) *Pyrococcus horikoshii* OT3, (17) *Aspergillus oryzae* RIB40, (18) *Yarrowia lipolytica*.

TABLE LEGENDS

Table 1: Data collection and refinement statistics (molecular replacement)

Values in parentheses represent data for the highest resolution shell. One crystal was used for each dataset. [a] $R_{sym} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I is the measured intensity of each reflection and <I> is the intensity from multiple observations of symmetry-related reflections. [b] R-factor $= \sum |F_{obs}| - |F_{calc}| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are the observed and the calculated structure factor amplitudes, respectively. [c] R_{free} is the R-factor calculated with reflections chosen at random and omitted from refinement.

Table 2: Mutations in BapA and their impact on precursor cleavage and substrate hydrolysis

[a] The activities of the purified enzymes were measured after 192h of incubation under standard assay conditions (see Methods section) and related to the activity of native BapA (100%) treated under the same conditions. [b] Due to the reduced or absent autoprocessing, the activities of the mutants E133A, S250A, S288A and E290A were measured from cell-free extracts containing the respective recombinantly expressed BapA Duet variants and related to the activity of cell-free extract containing native BapA Duet (100%).

Table 3: Kinetic analyses of the conversion of different β^3-homoamino-acid *p*-nitroanilides by the purified β-aminopeptidases BapA, DmpA and DmpA W137A

[a] The enzyme activities measured at pH 7.2 and 37°C showed a linear dependency on the substrate concentration between 0.1 and 5 mM. The k_{cat}/K_M values were calculated according

to the equation $k_{cat}/K_M = v/[Eo]*[S])$, where v is the rate of the reaction, [S] the concentration of the substrate and [Eo] the stoichiometric concentration of active centers (M_{BapA}: 38,610, M_{DmpA}: 40,440, $M_{DmpA\ W137A}$: 40,330). [b] For the calculation of the kinetic parameters, only the processed enzyme fraction in the protein solution was taken into account.

TABLES

Table 1: Data collection and refinement statistics (molecular replacement)

	Native BapA	***proBapA***
Data collection		
Space group	P21	P212121
Cell parameters: *a*, *b*, *c* (Å)	87.4, 96.7, 101.4, β = 108.2	99.7, 113.7, 126.0
Resolution range (Å)	48.8 – 1.45 (1.47 – 1.45)	29.4 – 1.8 (1.9 – 1.8)
Completeness (%)	98.5 (92.3)	99.1 (99.2)
Redundancy	3.24 (2.33)	4.40 (4.46)
R_{sym} [a]	4.2 (43.3)	7.1 (47.3)
$I / \sigma I$	17.0 (2.7)	15.6 (4.0)
Refinement		
Resolution (Å)	1.45	1.8
N° unique reflections	278534	131659
N° free reflections	6964	2634
R_{work} [b] / R_{free} [c]	13.5 / 15.3	15.5 / 18.0
N° of atoms		
Protein	10836	10639
Ligand / ions	83	44
Water	1741	1124
B-factors		
Protein	18.58	21.54
Ligand / ions	30.61	37.60
Water	31.06	31.79
R.m.s deviations		
Bond lengths (Å)	0.011	0.009
Bond angles (°)	1.445	1.138
Ramachandran		
Most favored (%)	1375 (96.6%)	1348 (96.9%)
Generously allowed (%)	44 (3.1%)	39 (2.8%)
Disallowed (%)	4 (0.3%)	4 (0.3%)

Values in parentheses represent data for the highest resolution shell. One crystal was used for each dataset. [a] $R_{sym} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I is the measured intensity of each reflection and $\langle I \rangle$ is the intensity from multiple observations of symmetry-related reflections. [b] R-factor $= \sum |F_{obs}| - |F_{calc}| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are the observed and the calculated structure factor amplitudes, respectively. [c] R_{free} is the R-factor calculated with reflections chosen at random and omitted from refinement.

Table 2: Mutations in BapA and their impact on precursor cleavage and substrate hydrolysis

Residue	***Point mutation***	***Precursor cleavage*** *(autoproteolysis)*	***Enzymatic activity*** *(catalysis)*	***Residue essential for***	
				autoproteolysis	catalysis
	wt	complete	100% [a]		
Glu133	E133A	slow	inactive [a,b]	no	yes
Lys248	K248A	slow	100% [a]	no	no
Asn249	N249A	slow	100% [a]	no	no
Ser250	S250A	no cleavage	inactive [b]	yes	yes
Ser288	S288A	no cleavage	<1% [b]	yes	yes
Glu290	E290A	no cleavage	24% [b]	yes	yes

[a] The activities of the purified enzymes were measured after 192h of incubation under standard assay conditions (see Methods section) and related to the activity of native BapA (100%) treated under the same conditions. [b] Due to the reduced or absent autoprocessing, the activities of the mutants E133A, S250A, S288A and E290A were measured from cell-free extracts containing the respective recombinantly expressed BapA Duet variants and related to the activity of cell-free extract containing native BapA Duet (100%).

Table 3: Kinetic analyses of the conversion of different β[3]-homoamino-acid *p*-nitroanilides by the purified β-aminopeptidases BapA, DmpA and DmpA W137A

Substrate	***BapA***			***DmpA***			***DmpA (W137A)*** [b]		
	K_M [mM]	V_{max} [µmol *min^{-1}mg^{-1}]	K_{cat}/K_M [M^{-1}s^{-1}]	K_M [mM]	V_{max} [µmol *min^{-1}mg^{-1}]	K_{cat}/K_M [M^{-1}s^{-1}]	K_M [mM]	V_{max} [µmol *min^{-1}mg^{-1}]	K_{cat}/K_M [M^{-1}s^{-1}]
H-βhGly-*p*NA	6.6 ±2.3	0.65 ±0.15	64 ±37	0.016 ±0.001	55.8 ±0.8	2,400,000 ±200,000	2.9 ±0.2	20.1 ±0.5	4600 ±400
H-β^{3}hAla-*p*NA	1.1 ±0.2	16.4 ±0.9	9600 ±1800	0.063 ±0.002	109 ±1	1,200,000 ±100,000	0.50 ±0.02	40.0 ±0.3	53,800 ±2,100
H-D-β^{3}hAla-*p*NA			< 10 [a]	17.0 ±2.7	1.3 ±0.2	52 ±15	0.38 ±0.13	0.017 ±0.002	30 ±13
H-β^{3}hPhe-*p*NA	3.5 ±1.3	79.3 ±14.5	14,500 ±7,800			<10 [a]	0.85 ±0.09	3.5 ±0.1	2,800 ±400
H-D-β^{3}hPhe-*p*NA			940 ±3 [a]			< 10 [a]	2.7 ±0.4	26.7 ±2.0	6,700 ±1600

[a] The enzyme activities measured at pH 7.2 and 37°C showed a linear dependency on the substrate concentration between 0.1 and 5 mM. The k_{cat}/K_M values were calculated according to the equation $k_{cat}/K_M = v/[Eo]*[S]$), where v is the rate of the reaction, [S] the concentration of the substrate and [Eo] the stoichiometric concentration of active centers (M_{BapA}: 38,610, M_{DmpA}: 40,440, $M_{DmpA\ W137A}$: 40,330). [b] For the calculation of the kinetic parameters, only the processed enzyme fraction in the protein solution was taken into account.

FIGURES

Figure 1A

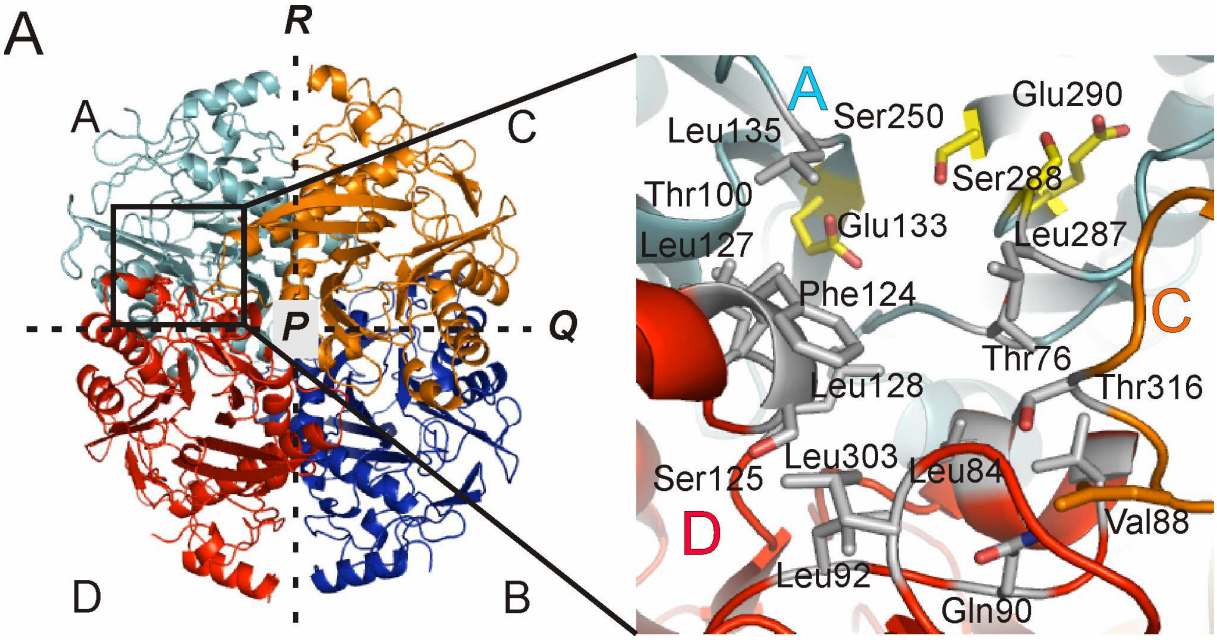


Figure 1B

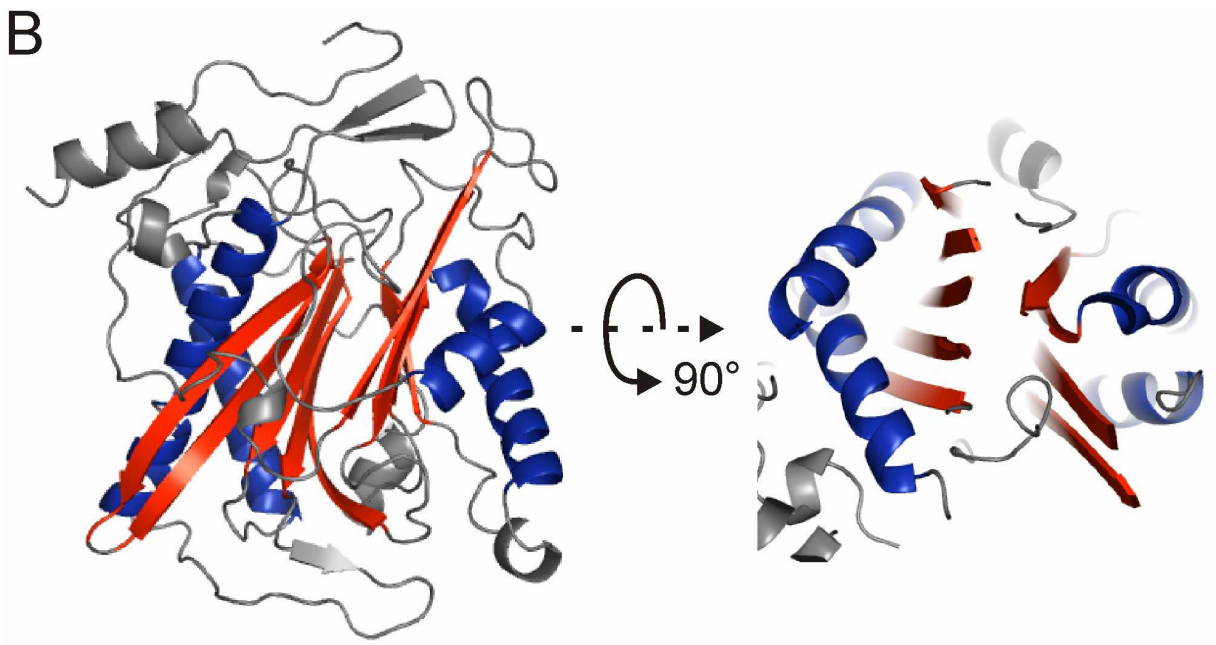


Figure 1C

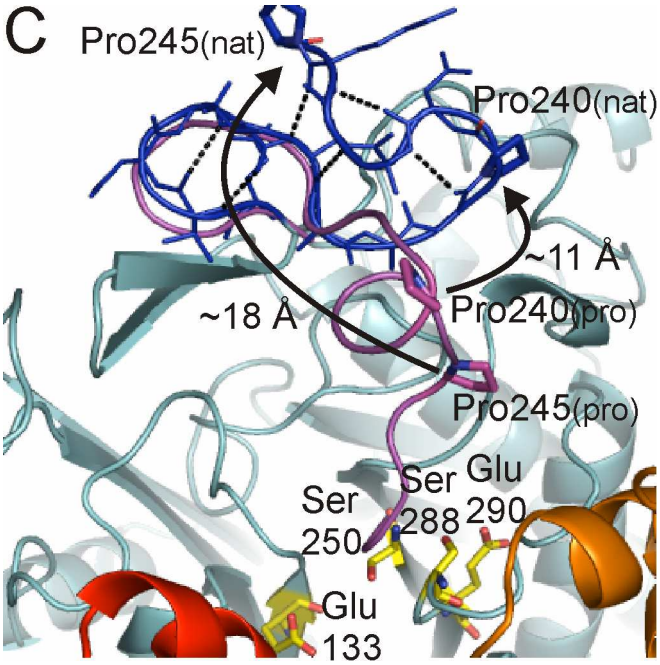

C
Pro245(nat)
Pro240(nat)
~11 Å
~18 Å
Pro240(pro)
Pro245(pro)
Ser
250
Ser
288
Glu
290
Glu
133

Figure 2A

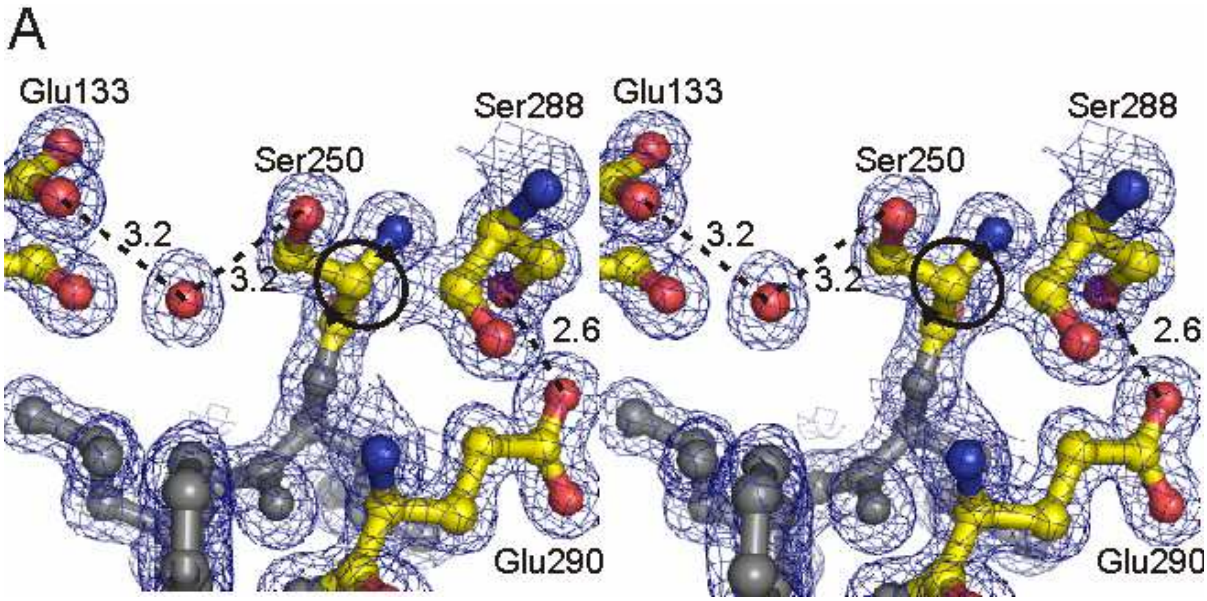


Figure 2B

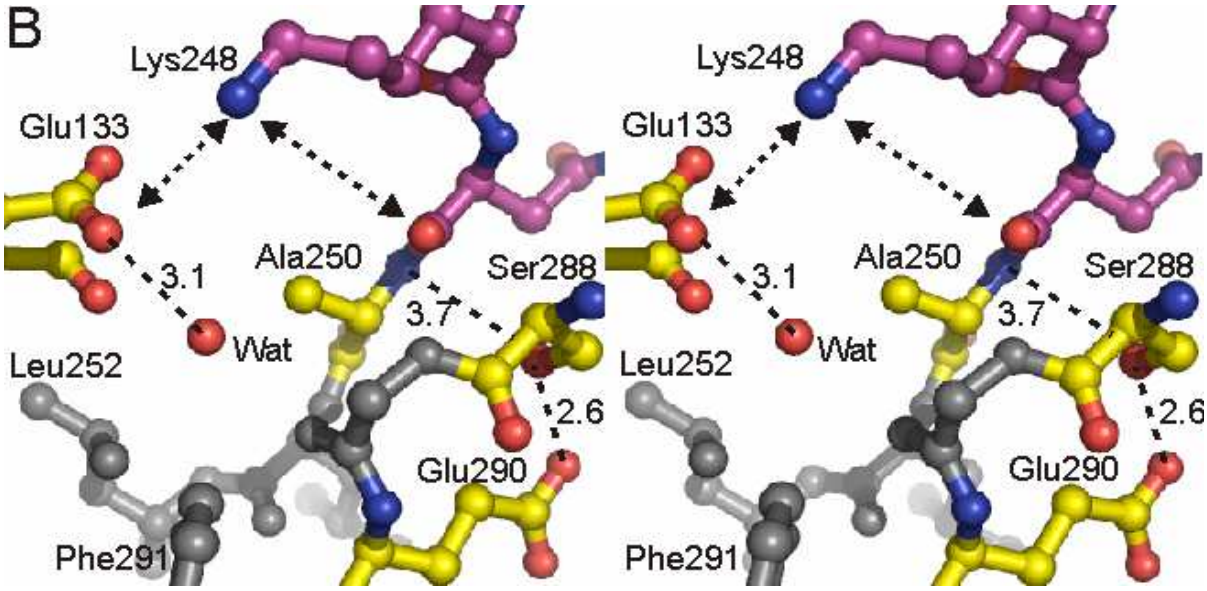


Figure 3

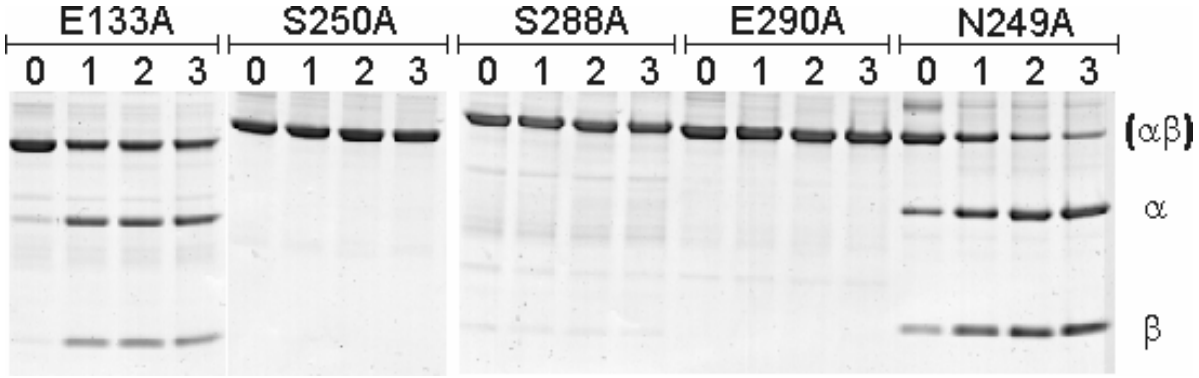


Figure 4A

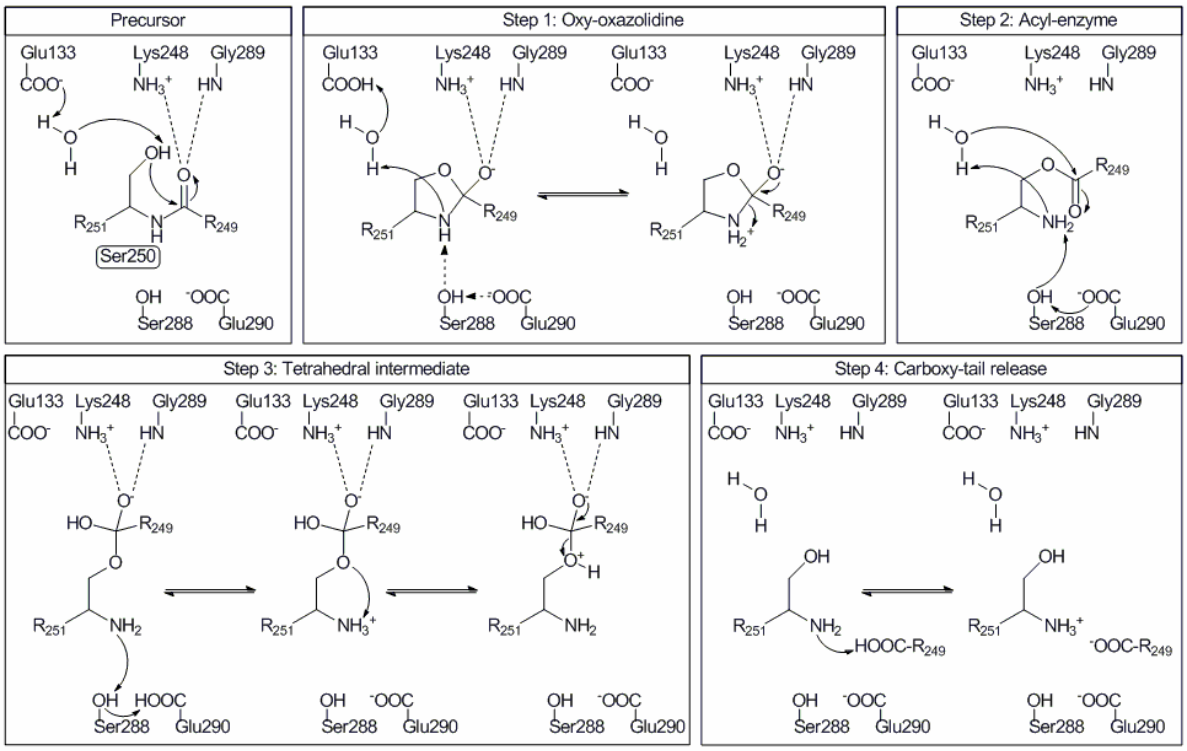


Figure 4B

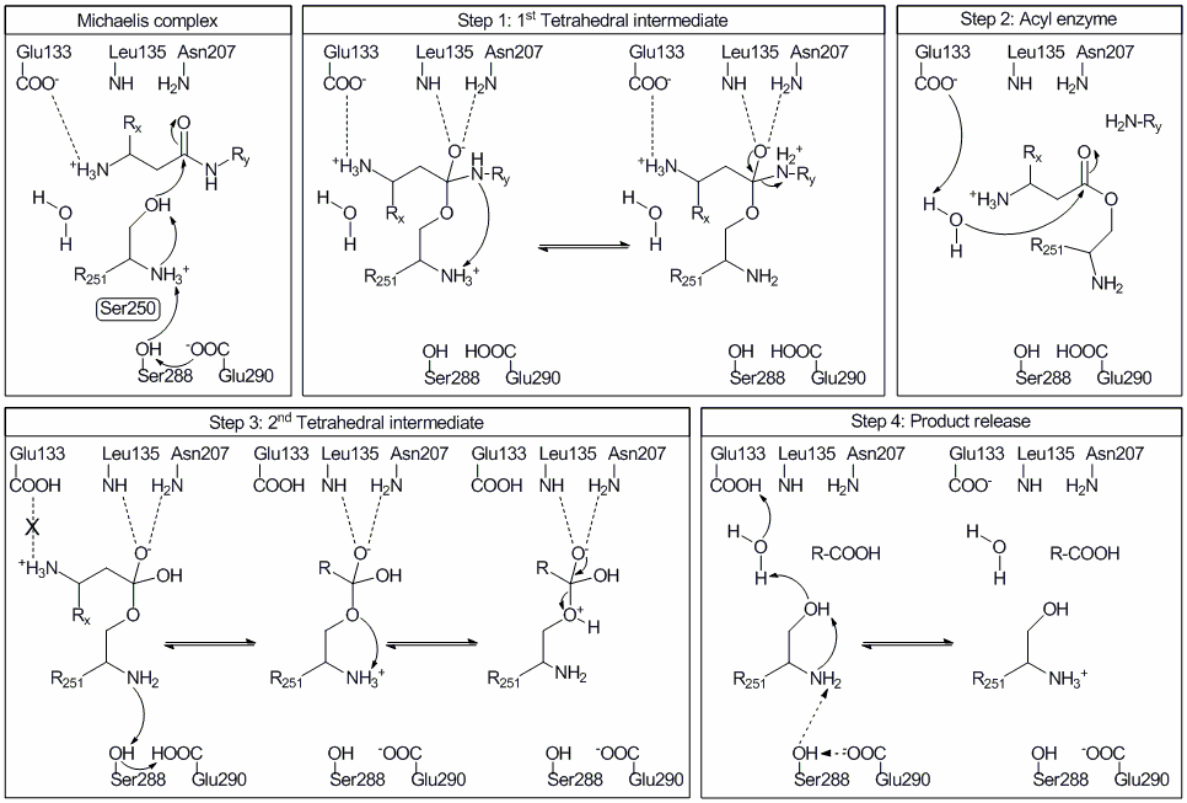
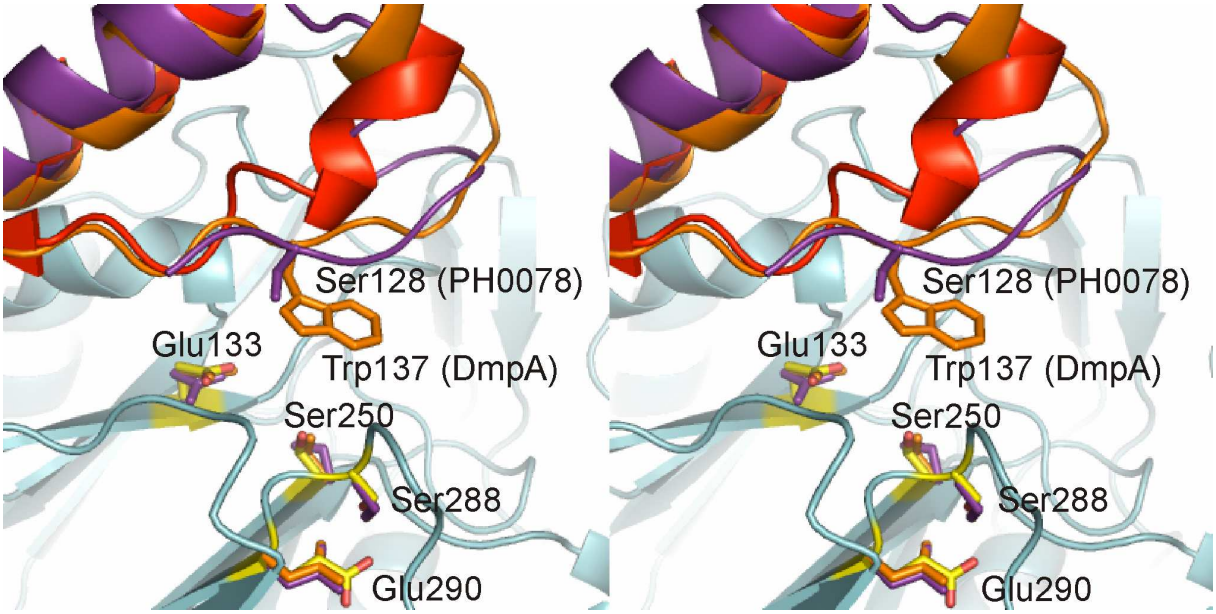


Figure 5


Ser128 (PH0078)
Glu133
Trp137 (DmpA)
Ser250
Ser288
Glu290
Ser128 (PH0078)
Glu133
Trp137 (DmpA)
Ser250
Ser288
Glu290

SUPPLEMENTAL INFORMATION

The supplemental information includes one figure showing the alternative catalytic mechanism (Figure S1) and sequences of primers for cloning and site-directed mutagenesis.

SUPPLEMENTARY FIGURE

Figure S1

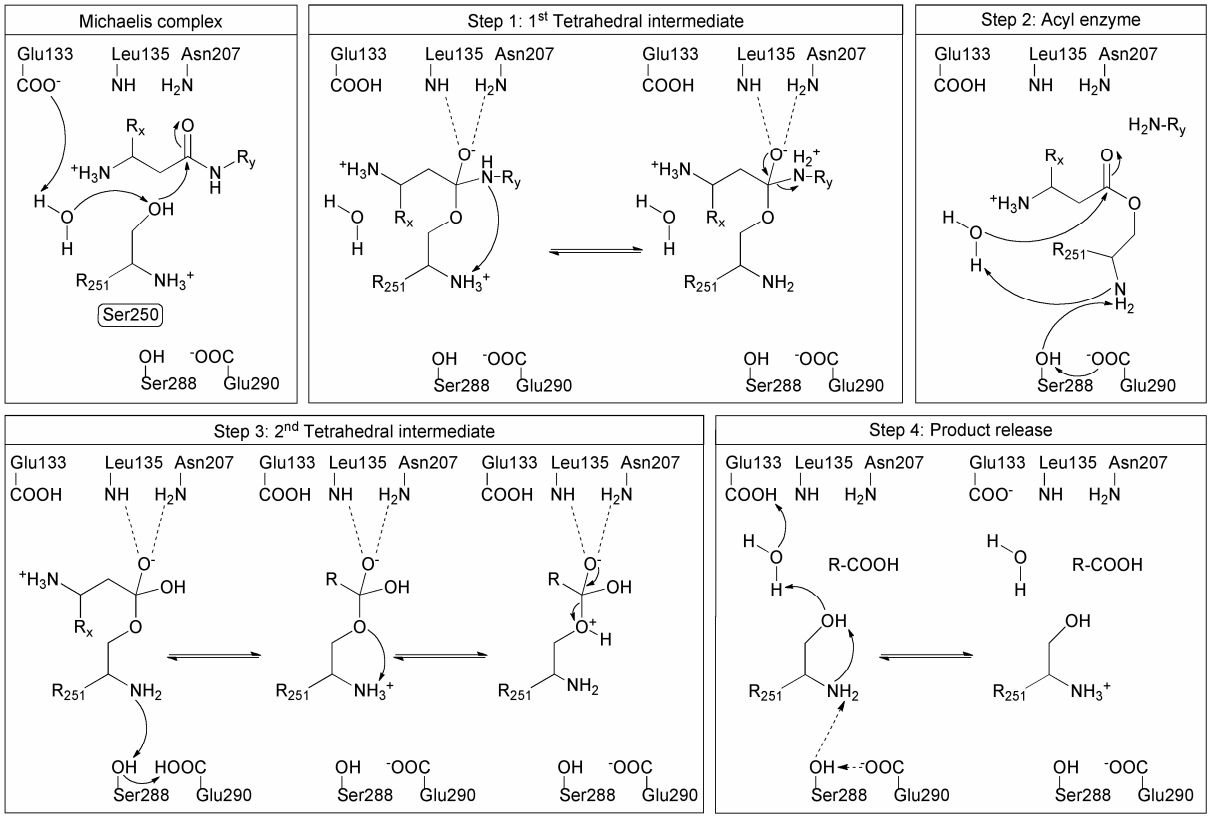


Figure S1: Schematic drawing of the alternative catalytic substrate-cleavage mechanism

Michaelis complex: Formation of a salt bridge between Glu133 and the substrate`s amino terminus. Activation of Ser250-Oγ by Glu133 and the catalytic water molecule, and initiation of the nucleophilic attack on the carbonyl group of the substrate. Step 1: Stabilization of the oxyanion by the backbone amide group of Leu135 and the δNH_2-group of Asn207. Substrate cleavage and formation of the acyl enzyme through proton transfers within the tetrahedral

intermediate. Step 2: Activation of the catalytic water molecule by Glu290 and initiation of the acyl-enzyme hydrolysis; Ser288 and the amino group of Ser250 enable the proton transfer. Step 3: Indirect protonation of the amino group of Ser250 by Glu290 and Ser288 and rupture of the tetrahedral intermediate through subsequent proton and electron transfer reactions. Step 4: Release of the cleaved product and reconstitution of the active-site residues for another catalytic cycle.

Solid arrows denote electron-pair shifts, dashed arrows denote catalytic influence; hydrogen bonds are indicated by dotted lines. R_{249} = Asn249, R_{251} = Leu251, R = *N*-terminal amino-acid residue of the substrate, Rx = side chain of the *N*-terminal amino acid, Ry = C-terminal substrate moiety.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Sequences of primers for cloning and site-directed mutagenesis were as follows:

bapA α-chain: fwd 5`-GGAATTCCATGGGGCCGCGCGCTCGCGATCT-3`,

bapA α-chain: rev 5`-GGAATTGGATCCTAATTCTTGTCCTGCGGCTTGCCT-3`,

bapA β-chain: fwd 5`-GGAATTCCCATATGTCGCTGCTGATCGTGATCGCT-3`

bapA β-chain: rev 5`-GGAATTCTCGAGTCACCGGCGCGGAAACCGCGCCT-3`

The designated point mutants were generated by site-directed mutagenesis using the QuikChange multi-site-directed mutagenesis kit (Stratagene, La Jolla, USA) with the following primers:

BapA S250A: 5`-CCGCAGGACAAGAATGCGCTGCTGATCGTG-3`

BapA S250C: 5`-CCGCAGGACAAGAATTGCCTGCTGATCGTG-3`

BapA E133A: 5`-CCGGTGGTCGCCGCAACGCTCGACAAC-3`

BapA N249A: 5`-CAAGCCGCAGGACAAGGCTTCGCTGCTGATCGTG-3`

BapA S288A: 5`-GCGGGGGCGCTTGCGGGTGAGTTCGCG-3`

BapA E290A: 5`-GCGCTTTCGGGTGCGTTCGCGCTCGCC-3`

BapA Duet S250A: 5`-GGAGATATACATATGGCGCTGCTGATCGTG-3`

BapA Duet S250C: 5`-GAAGGAGATATACATATGTGCCTGCTGATCGTGATC-3`

BapA Duet E133A: 5`-CCGGTGGTCGCCGCAACGCTCGACAAC-3`

BapA Duet S288A: 5`-GCGGGGGCGCTTGCGGGTGAGTTCGCG-3`

BapA Duet E290A: 5`-GCGCTTTCGGGTGCGTTCGCGCTCGCC-3`

Crystal Structures of BapA Complexes with β-Lactam-Derived Inhibitors Illustrate Substrate Specificity and Enantioselectivity of β-Aminopeptidases

(Manuscript 2)

Tobias Heck,[1] Tobias Merz,[2] Artur Reimer,[1] Dieter Seebach,[3] Daniel Rentsch,[4] Markus G. Grütter,[2] Hans-Peter E. Kohler,[1,§] and Birgit Geueke[1]

[1] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Department of Environmental Microbiology
Überlandstrasse 133, 8600 Dübendorf, Switzerland

[2] Biochemistry Institute, University of Zürich
Winterthurerstrasse 190, 8057 Zürich, Switzerland

[3] Laboratory of Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich
Hönggerberg HCI, Wolfgang-Pauli-Strasse 10, 8093 Zürich, Switzerland

[4] Empa, Swiss Federal Laboratories for Materials Science and Technology
Überlandstrasse 129, 8600 Dübendorf, Switzerland

[§] Corresponding author: Dr. Hans-Peter E. Kohler; Tel: +41 (0)44 8235521; Fax: +41 (0)44 8235547; e-mail: kohler@eawag.ch

HIGHLIGHTS

- The β-aminopeptidase BapA is an Ntn hydrolase with β-lactam binding properties
- Ampicillin and ampicillin-derived penicilloic acid are competitive inhibitors of BapA
- Salt bridge formation to BapA Glu133 is essential for ligand binding and positioning
- Conformational flexibility of β-peptides is required for substrate conversion by BapA

RUNNING TITLE

Crystal structures of BapA-inhibitor complexes

SUMMARY

β-Aminopeptidases have exclusive biocatalytic potential because they react with peptides comprised of β-amino acids that serve as building blocks for the design of non-natural peptidomimetics. In the present study, we identified the β-lactam antibiotic ampicillin and the ampicillin-derived penicilloic acid as novel inhibitors of the β-aminopeptidase BapA from *Sphingosinicella xenopeptidilytica* (K_i values of 0.69 and 0.74 mM, respectively). We report crystal structures of BapA in non-covalent complexes with these inhibitors, and with the serine-protease inhibitor AEBSF. All three inhibitors showed similar binding characteristics: The aromatic part of the molecule extended into a hydrophobic binding pocket of the active site and the free amino group formed a crucial salt bridge with Glu133 of BapA. The exact position of the inhibitors and structural details of the ligand-binding pocket illustrate the substrate specificity and the enantioselectivity of BapA-catalyzed reactions.

INTRODUCTION

The incorporation of backbone-elongated β-amino acids with proteinogenic side chains (Seebach et al., 2004) (**Figure 1**) gives peptides that may form unique secondary structures and that are resistant to degradation by most proteolytic enzymes (Frackenpohl et al., 2001; Wiegand et al., 2002). These properties give rise to interesting new biomedical applications for β-peptides as bioactive α-peptide mimetics (Aguilar et al., 2007; Seebach and Gardiner, 2008; Steer et al., 2002). To date, five sequence-related enzymes have been biochemically and functionally characterized; they posses the unusual property of cleaving N-terminal β-amino-acid residues from linear β- and mixed β,α-peptides that are otherwise resistant to proteolytic breakdown (Fanuel et al., 1999; Fuchs et al., 2010; Geueke et al., 2005; Heck et al., 2006; Komeda and Asano, 2005). Due to their common catalytic activity these β-peptide-converting catalysts are referred to as β-aminopeptidases (Geueke and Kohler, 2007). The growing demand for β-amino-acid-based building blocks (Liljeblad and Kanerva, 2006) has led to the development of biocatalytic applications for β-aminopeptidases, such as the preparation of enantiopure β-amino acids and the synthesis of β- and mixed β,α-peptides (Heck et al., 2007; Heck et al., 2010a; Heck et al., 2010b; Heck et al., 2009).

Alignment searches revealed many sequences from bacterial and eukaryotic genomes that have similarity with the known β-aminopeptidases (Geueke and Kohler, 2007). Within these sequences, the catalytically relevant amino acids are highly conserved, indicating that the ability to process β-peptides is not restricted to the few described β-aminopeptidases (Merz et al., this issue). The large abundance of β-aminopeptidase-like sequences suggests that these enzymes have important cellular functions that are as yet unknown. Although peptides solely composed of linear β-amino-acid residues do not occur in nature, many β-amino acids are

constituents of a wide variety of different, often highly bioactive substances, such as microcystins (Dawson, 1998; Luukkainen et al., 1993), taxol (Juaristi, 2005), bestatin (Suda et al., 1976; Umezawa et al., 1976), and L-carnosine (Bauer, 2005; Guiotto et al., 2005). β-Aminopeptidases may participate in the breakdown of such β-amino-acid-containing compounds to make them available for further metabolic conversions.

Antibiotics of the penicillin family constitute one particular group among the plethora of natural compounds that contain β-amino-acid substructures. They are of special interest for β-peptide research, because their characteristic β-lactam rings are cyclic derivatives of α,β-diamino-acid moieties (**Figure 2**). However, note that the antibiotic ampicillin, in addition to the β-lactam-containing core structure 6-aminopenicillanic acid (6-APA), also consists of the α-amino acid phenylglycine and may hence be viewed as a dipeptide with an N-terminal α-amino acid. The degradation of β-lactam-derived compounds is mainly catalyzed by two different classes of enzymes, namely by β-lactamases (EC 3.5.2.6) and by penicillin acylases (EC 3.5.1.11). β-Lactamases hydrolyze the amide bond of the β-lactam ring to a linear α,β-diamino-acid derivative (penicilloic acid) (Deshpande et al., 2004; Matagne et al., 1999). Penicilloic acids are usually antimicrobially inactive, but interestingly they act as competitive inhibitors of the β-lactamase-catalyzed reaction (Kiener and Waley, 1978). Several hundred β-lactamase variants, which are grouped into four different classes, have been described in the last decades (Bush and Jacoby, 2010; http://www.lahey.org/studies). Their genes are easily transferred between bacteria by horizontal gene transfer, causing the dissemination of resistance (Barlow, 2009). The second class of β-lactam-converting enzymes comprises penicillin acylases, which hydrolyze the amide bond of β-lactam antibiotics that links the β-

lactam core to a specific side chain (Chandel et al., 2008). Penicillin acylases are commercially important for the industrial chemo-enzymatic production of the β-lactam precursor 6-APA and of semisynthetic β-lactam antibiotics (Arroyo et al., 2003).

Although penicillin acylases and β-aminopeptidases do not share high sequence similarity, both enzyme classes possess structural and catalytic features of the N-terminal-hydrolase (Ntn) superfamily (Oinonen and Rouvinen, 2000). The Ntn hydrolases adopt a characteristic four-layered αββα-sandwich structure. One Ntn-hydrolase unit is formed by two polypeptide chains (α and β) that result from posttranslational autoproteolytic cleavage of an inactive precursor polypeptide (Artymiuk, 1995; Brannigan et al., 1995). Self-processing of the precursor is promoted by a conserved serine, cysteine or threonine residue, which thereupon becomes the catalytic N-terminal nucleophile of the newly formed β-polypeptide chain. Crystal structures of the β-aminopeptidases DmpA from *Ochrobactrum anthropi* LMG7991 (PDB ID: 1b65) (Bompard-Gilles et al., 2000b), BapA from *Sphingosinicella xenopeptidilytica* 3-2W4 (PDB ID: 3n2w), and of an immature BapA precursor mutant (PDB ID: 3n5i) (Merz et al., this issue) revealed that these enzymes are distinct from most other Ntn hydrolases by the orientation and connectivity of the secondary-structure elements; it is hence assumed that the topological and functional properties of Ntn hydrolases and of β-aminopeptidases result from convergent evolution of different evolutionary origins (Cheng and Grishin, 2005).

Based on point mutations at the BapA active site, we have devised a mechanism of the BapA-catalyzed reaction (Merz et al., this issue). Essential functions were assigned to the N-terminal catalytic nucleophile Ser250 that is activated by the dyad Ser288/Glu290, and to Glu133 that

forms a salt bridge with the N-terminal amino group of the peptide substrate. However, the molecular determinants of the unusual β-peptide specificity of β-aminopeptidases have not been investigated in detail, since all attempts to obtain crystal structures of β-aminopeptidase-ligand complexes and to identify β-aminopeptidase-specific inhibitors have as yet been unsuccessful. In the present study, we report the inhibition of BapA by the β-lactam antibiotic ampicillin and by an ampicillin-derived hydrolysis product. Furthermore, we present crystal structures of non-covalent BapA complexes with these inhibitors and with the broad-spectrum serine-protease inhibitor 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF, trade name: pefabloc SC), providing structural details of ligand binding and mechanistic details of β-peptide transformations catalyzed by β-aminopeptidases.

RESULTS AND DISCUSSION

Crystal structure of BapA in complex with AEBSF

Biochemical studies with the β-aminopeptidase BapA demonstrated the inhibition of the enzyme in the presence of high concentrations of the broad-spectrum serine-protease inhibitor AEBSF (Geueke et al., 2006). We determined the crystal structure of a non-covalent BapA-AEBSF complex (BapA-AEBSF) at 1.8 Å resolution; crystallographic data are summarized in **Table 1**. One molecule of AEBSF was found in each of the four active sites at the interface of three adjacent subunits of the BapA $(\alpha\beta)_4$-heterooctamer. The amino group of AEBSF formed a salt bridge with the carboxyl group of Glu133 and a hydrogen bond to a catalytically relevant water molecule (see Merz et al., this issue). Thr76, Thr100, Leu135, and Leu287 (e.g. subunit A) together with residues of the two neighboring αβ-subunits (e.g. subunit C: Thr316; subunit D: Leu84, Val88, Gln90, Leu92, Phe124, Ser125, Leu127, Leu128, and Leu303) constituted a hydrophobic pocket that supported ligand binding by *van der Waals* interactions (**Figure 3**). The aromatic ring of Phe124, which showed very weak electron density in the native structure of BapA (Merz et al., this issue), was well defined through a π-stacking interaction with the aromatic ring of AEBSF. In contrast to the proposed irreversible inhibition mechanism of AEBSF, which proceeds by formation of a covalent bond between a catalytically active nucleophile and the sulfonyl group of the inhibitor (Powers et al., 2002), the structure of the BapA-AEBSF complex showed binding of the inhibitor in a non-covalent manner. While the fluoro-sulfonyl terminus pointed away from the catalytic nucleophile Ser250-Oγ into the hydrophobic binding pocket, the amino group of AEBSF formed a salt bridge to Glu133.

Inhibition of the BapA by penicillin-derived β-lactam antibiotics

To investigate β-aminopeptidase-catalyzed biotransformations of cyclic β-amino acids and β-lactams, we examined the potential of the β-aminopeptidases BapA and DmpA to process the penicillin derivatives ampicillin, penicillin G and carbenicillin. These compounds contain a β-lactam ring as part of their common 6-APA core structure (**Figure 2**) and differ by their D-phenylglycyl, phenylacetyl and D-phenylmalonyl acyl groups, respectively. With neither enzyme we observed deacylation or hydrolysis of the β-lactam amide bond, demonstrating that both BapA and DmpA lack penicillin-acylase and β-lactamase activity with the tested lactams.

Furthermore, we tested the inhibitory effect of the three β-lactam antibiotics and of the ampicillin-derived penicilloic acid Amp_{hyd} (**Figure 2**) on the BapA- and DmpA-catalyzed conversion of the chromogenic reporter substrate H-β^3hAla-*p*NA. Despite the structural similarity of ampicillin, penicillin G and carbenicillin with respect to their 6-APA core structure, only ampicillin inhibited substrate conversion by BapA with a K_i value of 0.69 mM (**Table 2**). Ampicillin differs from penicillin G and carbenicillin by the presence of an α-amino group in its D-phenylglycyl moiety (**Figure 2**). With the penicilloic acid Amp_{hyd} a K_i value of 0.74 mM was observed; thus, the structural change associated with the conversion of ampicillin to Amp_{hyd}, *i.e.* hydrolytic opening of the β-lactam ring, did not affect the inhibitory effect on substrate conversion by BapA. In contrast to BapA, DmpA was not significantly inhibited by any of the three β-lactam derivatives or by the hydrolyzed species Amp_{hyd}.

Crystal structures of BapA in complex with ampicillin and Amp_{hyd}

Crystal structures of BapA in non-covalent complex with ampicillin (BapA-Amp) and with the (5*S*)-epimer of enzymatically produced Amp_{hyd} (BapA-Amp_{hyd}) at 1.7 Å and 1.85 Å resolution, respectively, were obtained by soaking BapA crystals with the respective compounds; crystallographic data are summarized in **Table 1**. Both ligands interacted with the active site of the enzyme in a similar manner, forming a dense hydrogen-bonding network. The α-amino group of the D-phenylglycyl moiety of ampicillin formed a salt bridge with Glu133 and a hydrogen bond with the catalytically active water molecule (**Figure 4A**). As observed for the aromatic ring of AEBSF in the structure of the BapA-AEBSF complex, the phenyl ring of ampicillin extended into the hydrophobic pocket of the BapA active site (*cf.* **Figure 3** and **Figure 4**). The backbone nitrogen of Leu135 and the carbonyl oxygen of Leu287 formed hydrogen bonds with the amide bond that links the acyl group to the 6-APA core of ampicillin. Furthermore, the carboxyl group attached to the thiazolidine ring formed a salt bridge with the side chain of Arg138. In the case of Amp_{hyd}, the main interactions between the D-phenylglycyl acyl group and the active site of BapA were similar to those observed for binding of ampicillin (*cf.* **Figure 4A** and **4B**). Because of the missing β-lactam substructure, the core of Amp_{hyd} is more relaxed than that of ampicillin and the position of the thiazolidine is less rigid. Due to this structural difference, the interaction of the carboxyl group attached to the thiazolidine ring with Arg138 was not present in the BapA-Amp_{hyd} complex. Instead, the additional carboxyl group of Amp_{hyd} was in close proximity of Asn207-NH_2 and the amino terminus of Ser250, compensating for the missing salt bridge with Arg138.

Models of tetrahedral BapA-substrate complexes

Based on structural similarities of the N-terminal moiety of Amp_{hyd} to a linear α-peptide, we modeled β^3-homophenylalanine residues of L- and D-configuration onto the backbone of the ligand in the BapA-Amp_{hyd} complex. For model building, we positioned the carbonyl group of the substrate into appropriate distance (1.6 Å) and trajectory (105 ± 5°) (Bürgi and Dunitz, 1983; Bürgi et al., 1974) for nucleophilic attack by Ser250-Oγ, and aimed for the best possible match of the β^3-homophenylalanine residues on the electron density of Amp_{hyd}. Tetrahedral reaction intermediates were formed by the addition of Ser250-Oγ to the β^3-homophenylalanine residues. Rotation around the Cα-Cβ bond of the β-amino-acid backbone gives rise to two synclinal conformers ((+)-sc and (-)-sc) and an antiperiplanar conformer (ap) for a β-amino-acid residue (Seebach et al., 2004). It turned out that the (+)-sc conformer of L-β^3-homophenylalanine and the (-)-sc conformer of D-β^3-homophenylalanine represents the only energetically favorable conformation of the respective compound that met the steric demands for successful formation of a tetrahedral enzyme-substrate intermediate (**Figure 5A** and **Figure S1**). Complex formation with the other conformers ((-)-sc and ap of L-β^3-homophenylalanine, (+)-sc and ap of D-β^3-homophenylalanine) could be excluded due to massive steric repulsions between the side chain of the β-amino-acid residue and residues of the BapA active site (**Table 3** and **Figure S1**). In the favored conformations, the hydrophobic side chains and the amino groups of the covalently bound L- and D-β^3-homophenylalanine residues matched well with the positions of the ligands in the BapA-AEBSF, BapA-Amp_{hyd}, and BapA-Amp complexes (**Figure 5A** and **5B**). The proximity of Asn207-NH_2 and Leu135-NH to the oxyanion of the modeled tetrahedral intermediates suggested that these residues stabilize the negatively charged tetrahedral intermediate during the catalytic reaction (for mechanistic details see Merz et al., this issue).

Implications of ligand binding for β-peptide conversions by BapA and DmpA

The structures of the BapA-inhibitor complexes showed a good overlap of the amino group and the aromatic ring of AEBSF with the positions of the phenylglycine acyl groups of ampicillin and Amp_{hyd} (**Figure 5**). In combination with previous information on the substrate specificity of BapA for N-terminally unprotected β-peptides (Geueke et al., 2006; Heck et al., 2006) and results obtained from a mutational analysis of BapA (Merz et al., this issue), the crystal structures of the BapA-ligand complexes (**Figure 3** and **Figure 4**) and the inhibition of BapA by ampicillin-derived compounds (**Table 2**) provide strong evidence that the presence of a free amino group plays a key role for ligand recognition and binding by BapA. Due to their amino groups attached to aromatic rings through one or two CH_2-groups, the β-lactam-derived ligands (ampicillin and Amp_{hyd}) and AEBSF have structural similarity to β^3-homophenylglycine and β^3-homophenylalanine residues, respectively (*cf.* **Figure 2** and **Figure 1**). Therefore, the formation of non-covalent complexes between BapA and these ligands suggests that the binding of β-peptidic substrates is promoted by similar interactions within the BapA active site, *i.e.* (*i*) salt-bridge formation between Glu133 and the amino group of the substrate's N-terminal β-amino-acid residue, and (*ii*) hydrophobic interactions between the ligand-binding pocket of BapA and the side chain of the substrate.

Based on the models of the tetrahedral BapA-substrate intermediates (**Figure 5A**), we argue that the exclusive specificity of the enzyme for peptides with backbone-elongated N-terminal β-amino-acid residues is governed by the molecular geometry of the active site, in particular by the orientations and distances of the substrates' functional groups to Glu133 and to Ser250-Oγ. Our results show that compounds, such as ampicillin and Amp_{hyd}, mimicking the N-termini of conventional α-peptides bind to the active site of BapA, but are not converted.

Hence, it seems that the lack of α-aminopeptidase activity of BapA is caused by the "rigidity" of the α-amino-acid backbone and by an insufficient distance between the functional groups (NH_2 and CO); these limitations prevent appropriate positioning of the carbonyl group for nucleophilic attack under the restrictions given by the BapA active site. The extended backbone of β-amino-acid residues confers a degree of rotational flexibility around the Cα-Cβ bond, which is not present in α-amino-acid residues; due to the increase in backbone length and the additional degree of rotational freedom, a β-amino-acid residue is able to adopt a stable conformation allowing for nucleophilic attack by Ser250-Oγ, while maintaining the salt bridge to Glu133 as well as the stabilization of the tetrahedral intermediate by the Leu135/Asn207 oxyanion hole.

In contrast to BapA, which exclusively cleaves peptides carrying N-terminal β-amino acids with aliphatic, aromatic or functionalized side chains, DmpA preferentially reacts with substrates carrying unsubstituted or sterically undemanding N-terminal β-amino acids (e.g. β-homoglycine and β^3-homoalanine), but also α-amino acids (Fanuel et al., 1999; Geueke et al., 2006; Heck et al., 2006; Heck et al., 2010b). The superposition of DmpA on the modeled BapA-substrate intermediate perfectly illustrates this steric restriction of the DmpA active site (**Figure 5B**). While BapA has a rather open ligand-binding pocket, the active site of DmpA is largely occluded by a loop ranging from Gln131 to Trp137; in particular, the side chain of Trp137 protrudes far towards the catalytic nucleophile Ser250 and could hence (*i*) force substrates with N-terminal α-amino acids into a position that allows for nucleophilic attack of Ser250-Oγ, and (*ii*) restrict catalysis by DmpA to substrates with sterically undemanding side chains. In fact, the preference of DmpA to process substrates with small N-terminal β-amino acids could be changed towards substrates with larger side chains by the single amino-acid

mutation of Trp137 to alanine (Merz et al., this issue). Furthermore, we hypothesize that the bulky side chain of Trp137 also prevents the inhibition of DmpA by ampicillin or Amp_{hyd}, because both carry aromatic rings as part of their phenylglycyl moieties that are prone to clash with the side chain of Trp137 (**Figure 5B**).

Enantioselectivity of the BapA-catalyzed reactions

The BapA-catalyzed kinetic resolution of racemic β^3-amino-acid amides with different aliphatic side chains has recently been demonstrated to proceed with high L-enantioselectivity (Heck et al., 2009). As shown in the models presented in **Figure 5A**, tetrahedral reaction intermediates could be formed between Ser250-Oγ of BapA and β^3-amino-acid residues of both L- and D-configuration. Our models indicate that the distance between Glu133-COO^- and the substrate's amino terminus is larger for the complex with the (-)-sc conformation than for the one with the (+)-sc conformation of the substrate molecules, indicating a negative effect on the strength of the salt bridge for the former complex. Therefore, we propose that in the active site of BapA the (+)-sc conformation of an L-β^3-amino-acid residue is energetically favored over the (-)-sc conformation of a D-β^3-amino-acid residue. The enantiomeric excess of the L-β^3-amino-acid product formed by the BapA-catalyzed kinetic resolution of racemic β^3-amino-acid amides also depends on the size of the substrate's side chain; substrates with small aliphatic substituents are more efficiently resolved by BapA than substrates with sterically demanding side chains (Heck et al., 2009). We propose that the correlation between decrease of enantioselectivity and increase of side-chain size of the N-terminal β-amino-acid residue is due to additional non-polar interactions of the hydrophobic ligand-binding pocket with large side chains of the substrate. Such hydrophobic effects will improve the stability of

complexes formed between BapA and large D-β^3-amino-acid residues of the (-)-sc conformation and partially compensate for a weak salt bridge to Glu133.

Besides its high catalytic activity with substrates composed of L-β^3-amino acids, BapA slowly processes a diastereomeric mixture of the β^2-dipeptide H-β^2hPhe-β^2hAla-OH with high D-enantioselectivity (Heck et al., 2010b). In order to elucidate the structural details of this enantioselective reaction, we modeled the reaction intermediates formed by the nucleophilic addition of BapA Ser250-Oγ to the carbonyl group of the (+)-sc conformers of D- and L-β^2-homophenylalanine residues (**Figure S2**). Due to the previously mentioned flexibility of the Phe124 side chain, the tetrahedral reaction intermediate of D-β^2-homophenylalanine could be placed in the active site of BapA. In contrast, reasonable placing of the tetrahedral reaction intermediate of the corresponding L-enantiomer was not possible because of a steric conflict between Cγ of the substrate's side chain and Thr76 of BapA. This fully explains the high enantioselectivity towards these substrates.

β^2-Dipeptides with small and therefore less demanding side chains at the N-terminal amino-acid residue are not converted by BapA, irrespective of their configuration(Heck et al., 2010b). As suggested for the conversion of substrates carrying N-terminal D-β^3-amino-acid residues with side chains of different sizes, the preference of BapA for bulky β^2-amino-acid residues likewise can be explained by the stabilization of the enzyme-substrate complex by increased hydrophobic interactions between large amino-acid side chains and the ligand-binding pocket.

Physiological significance of ampicillin binding and β-peptide conversion by BapA and other β-aminopeptidases

The position and the orientation of ampicillin in the crystal structure of the BapA-Amp complex (**Figure 4**) explain why BapA exhibits neither penicillin-acylase nor β-lactamase activity. Most importantly, the observed distance of Ser250-Oγ to the carbonyl-carbon atom of the phenylglycyl moieties (2.6 Å) or to the one of the β-lactam ring (4.3 to 4.6 Å) was too large to allow a nucleophilic attack. Moreover, the observed angle between Ser250-Oγ and the carbonyl plane of the phenylglycyl acyl group (84.9 to 88.5°) and the one between Ser250-Oγ and the carbonyl plane of the β-lactam (81.4 to 83.5°) was not in accordance with the required trajectory of 105 ± 5°, at which nucleophilic attack of a carbonyl carbon may occur (Bürgi and Dunitz, 1983; Bürgi et al., 1974). Based on these structural and biochemical observations, we have rationalized why BapA is an ampicillin-binding, but not an ampicillin-converting enzyme. Although the present study outlines the structural determinants of the exclusive β-peptide specificity of BapA, the natural substrates of this enzyme and of other β-aminopeptidases as well as the physiological significance of ampicillin recognition and inhibition are as yet unknown and remain to be unraveled. The D-aminopeptidase (DAP) from *O. anthropi*, another well-studied example of an aminopeptidase with an unusual substrate specificity that is inhibited by ampicillin and other β-lactam-derived compounds (Asano et al., 1992), exhibits sequence and structural similarities to β-lactamases and D,D-carboxypeptidases, and was classified as a new member of the penicillin-recognizing enzyme family (Bompard-Gilles et al., 2000a). DAP could be converted to an enzyme with D,D-carboxypeptidase activity and new penicillin-binding properties by directed mutagenesis (Delmarcelle et al., 2005). Although there are no obvious sequence similarities between BapA and β-lactam-converting or -binding enzymes, future investigations will need to address the

catalytic potential of the many uncharacterized proteins that were retrieved from sequencing projects and that are similar to the characterized β-aminopeptidases (Merz et al., this issue) towards transformations of β-lactams.

EXPERIMENTAL PROCEDURES

General remarks

AEBSF and the β-lactam antibiotics ampicillin, penicillin G, and carbenicillin were purchased from Sigma-Aldrich (Buchs, Switzerland). H-β^3hAla-*p*NA was prepared chemically following a published procedure (Heck et al., 2006). The enzymatic conversions of the β-lactams were followed by HPLC analysis on a Dionex HPLC system equipped with a P680 pump, an ASI-100 automated sample injector, an UltiMate 3000 thermostatted column compartment, and a UVD 340U photodiode array detector (Dionex, Sunnyvale, CA, USA). Samples (10 µl) were analyzed on a reversed-phase HPLC stationary phase Nucleodur C18-Pyramid (250 × 4 µm, 5 µm particle size; Macherey-Nagel, Düren, Germany) that was equilibrated with 0.1% formic acid in water for 3 min at a column temperature of 20°C. The compounds were separated with a linear gradient of 0 to 60% acetonitrile within 15 min at a constant flow rate of 1 ml/min and detected at a wavelength of 205 nm. Mass spectra of all compounds were recorded with an API 4000 liquid chromatography/tandem MS system connected to an Agilent 1100 LC system. For protein determination, we used five-fold concentrated Bradford reagent (Bio-Rad, Rheinach, Switzerland) and bovine serum albumin (BSA) as a standard; absorbance measurements were performed at 595 nm with a Specord S 100 spectrophotometer (Analytik Jena, Jena, Germany). Proteins were analyzed by SDS-PAGE using pre-cast 10% Novex tricine gels (Invitrogen AG, Basel, Switzerland) according to the manufacturer's instructions. The protein gels were stained with Coomassie Brillant Blue and the purities of the enzymes were estimated by calculating the relative intensities of the protein bands with a GS-800 calibrated imaging densitometer and the software Quantity One (Bio-Rad, Rheinach, Switzerland).

Enzymatic production of Amp_{hyd}

Amp_{hyd} was produced from ampicillin by enzymatic conversion with the β-lactamase that was encoded on the pET3c plasmid (Novagen, Madison, WI, USA), constitutively expressed in *E. coli* BL21(DE3) pLysS and partially purified by anion-exchange and hydrophobic-interaction chromatography. Ampicillin (50 mg) was dissolved in H_2O (5.2 ml) and the enzymatic reaction was started by the addition of the β-lactamase-containing solution to yield a final protein concentration of 15 μg/ml. After 16 h of incubation at 30°C, the enzymatic reaction was stopped by removing all proteins > 10 kDa by centrifugation with Centricon YM-10 centrifugal devices (10 kDa MWCO; Millipore Corp., Billerica, CA, USA). The formation of Amp_{hyd} was verified by HPLC-MS analysis (t_R = 10.3 min, $[M + H]^+$: 368.0 (calcd. *M*: 367.1), purity after peak integration > 90%). NMR analysis showed that the primary reaction product (5*R*)-Amp_{hyd} undergoes epimerization in aqueous solution at C-5 to form (5*S*)-Amp_{hyd} (see Supplemental Information available online). The solution containing Amp_{hyd} was frozen to -80°C and freeze-dried.

Expression and purification of BapA for protein crystallization

The β-aminopeptidases BapA from *S. xenopeptidilytica* 3-2W4 was expressed without its N-terminal 29-amino-acid signal sequence in a pET3c-expression system, purified from the *E. coli* host (Geueke et al., 2006) and crystallized (Merz et al., this issue). BapA crystals were soaked for 5 min in a solution of 1.5 M ammonium sulfate, 100 mM HEPES (pH 7.5) containing 50 mM freshly dissolved AEBSF and for 30 min in 1.5 M ammonium sulfate, 100 mM HEPES (pH 7.5) containing saturating concentrations of ampicillin or Amp_{hyd}. After

cryo-protection in a solution containing 1.5 M ammonium sulfate, 100 mM HEPES (pH 7.5) and 30% glycerol, the crystals were flash-frozen in liquid nitrogen.

Structure determination

Data were collected at the Swiss Light Source beamline PX on a 6M Pilatus detector in a cryostream at 100 K at 1 Å wavelength. Diffraction data on a total of 720 frames were recorded with an oscillation range of 0.5° per frame. The data was processed with the program XDS (Kabsch, 1993). Initial phases were obtained by molecular replacement using the program PHASER (McCoy et al., 2007) and the structure of BapA (PDB ID: 3n2w) as search model. The asymmetric units each contained one molecule of the active BapA $(\alpha\beta)_4$-heterooctamer. The final model of BapA-AEBSF contained residues 1-245 / 250-371 for chain A, 1-245 / 250-317 for chain B, 1-245 / 250-371 for chain C, and 1-245 / 250-371 for chain D. The final model of BapA-Amp contained residues 1-245 / 250-371 for chain A, 1-243 / 250-317 for chain B, 1-244 / 250-371 for chain C, and 1-245 / 250-371 for chain D. The final model of BapA-Amp_{hyd} contained residues 1-245 / 250-371 for chain A, 1-242 / 250-317 for chain B, 1-244 / 250-371 for chain C, and 1-244 / 250-371 for chain D. The refinement of the structures was carried out with the program PHENIX (Adams et al., 2002). Model building and superpositions for figures were performed with the programs COOT (Emsley and Cowtan, 2004) and SSM (Krissinel and Henrick, 2004), respectively. Figures were created with the program Pymol (De Lano, 2002).

Expression of BapA and DmpA for enzymatic activity assays

To exclude the possibility of β-lactamase cross-contamination from enzyme expression in a β-lactamase-encoding pET3c expression system, the genes *bapA* from *S. xenopeptidilytica* 3-2W4 and *dmpA* from *O. anthropi* LMG7991 were cloned into pET9c vectors yielding pAR116 and pAR114, respectively (Heck et al., 2006; Heck et al., 2010b). The enzymes were recombinantly expressed in *E. coli* BL21(DE3) pLysS and purified according to established procedures (Geueke et al., 2006; Heck et al., 2006). The lyophilized enzyme powders were dissolved in a 10 mM potassium-phosphate buffer at pH 7.2, and the protein contents of the enzyme stocks were determined spectrophotometrically.

Penicillin-acylase and β-lactamase activity assay

The reaction mixtures contained 1 mM of the β-lactam antibiotics ampicillin, penicillin G or carbenicillin in 100 mM potassium-phosphate buffer at pH 7.2 and 37°C. The enzymatic reactions were initiated by the addition of one of the enzymes BapA or DmpA. Samples were withdrawn at intervals over 24 h from the reaction mixtures and the enzymatic reactions were terminated by heating at 90°C for 3 min. The samples were analyzed by reversed-phase HPLC and HPLC-MS.

Enzyme-inhibition assay

The inhibitory effects of the β-lactam antibiotics ampicillin, penicillin G or carbenicillin and of the penicilloic acid Amp_{hyd} on the β-aminopeptidase-catalyzed conversion of the reporter substrate H-β^3hAla-*p*NA were determined spectrophotometrically (Merz et al., this issue). The

assay mixtures contained the reporter substrate and the inhibitors in various molar ratios. Experimental data was fitted to the competitive inhibition model (Eq. 1) by non-linear regression analysis with the VisualEnzymics software (Softzymics, Princeton, NJ, USA) for the program IGOR Pro (WaveMetrics, Oswego, OR, USA).

$$v = \frac{V_{\max} \cdot [S]}{K_m \cdot (1 + \frac{[I]}{K_i}) + [S]} \qquad \text{(Eq. 1)}$$

PDB ACCESSION NUMBERS

BapA in complex with AEBSF (BapA-AEBSF, PDB ID: 3n33); BapA in complex with ampicillin (BapA-Amp, PDB ID: 3ndv); BapA in complex with Amp_{hyd} (BapA-Amp_{hyd}, PDB ID: 3nfb).

ACKNOWLDEDGEMENTS

We thank the staff of beamline PX of the Swiss Synchrotron Light Source (PSI) in Villigen, Switzerland for excellent technical assistance. This work was supported by the Deutsche Bundesstiftung Umwelt (T.H., DBU Project No. 13176-32) and the Swiss National Science Foundation (B.G., SNF Project No. 3152A0-100770). T.M. was supported by the Swiss National Science Foundation (grants to M.G.G.) and the Swiss NCCR structural biology program.

AUTHOR CONTRIBUTIONS

H.P.K. and B.G. conceived the work. T.H. and B.G. carried out protein expression, purification and crystallization experiments. T.H. and A.R. carried out kinetic analyses. T.M. carried out crystallographic data collection, structure determination, refinement and modeling. M.G.G. assisted the structure determination. D.R. carried out the NMR analysis of the inhibitors. T.H. and T.M. wrote the paper and prepared the figures. D.S. contributed to the β-peptide substrates. D.S., M.G.G., B.G. and H.P.K. contributed to discussion, data interpretation, and manuscript preparation. M.G.G., B.G. and H.P.K. supervised the work.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr., Sect. D: Biol. Crystallogr. *58*, 1948-1954.

Aguilar, M.-I., Purcell, A. W., Devi, R., Lew, R., Rossjohn, J., Smith, A. I., and Perlmutter, P. (2007). β-Amino acid-containing hybrid peptides – new opportunities in peptidomimetics. Org. Biomol. Chem. *5*, 2884-2890.

Arroyo, M., de la Mata, I., Acebal, C., and Castillon, M. P. (2003). Biotechnological applications of penicillin acylases: state-of-the-art. Appl. Microbiol. Biotechnol. *60*, 507-514.

Artymiuk, P. J. (1995). A sting in the (N-terminal) tail. Nat. Struct. Biol. *2*, 1035-1037.

Asano, Y., Kato, Y., Yamada, A., and Kondo, K. (1992). Structural similarity of D-aminopeptidase to carboxypeptidase DD and β-lactamases. Biochemistry *31*, 2316-2328.

Barlow, M. (2009). What antimicrobial resistance has taught us about horizontal gene transfer. In Horizontal gene transfer: genomes in flux, M. B. Gogarten, J. P. Gogarten, and L. Olendzenski, eds. (Humana Press), pp. 397-411.

Bauer, K. (2005). Carnosine and homocarnosine, the forgotten, enigmatic peptides of the brain. Neurochem. Res. *30*, 1339-1345.

Bompard-Gilles, C., Remaut, H., Villeret, V., Prangé, T., Fanuel, L., Demarcelle, M., Joris, B., Frère, J.-M., and Van Beeumen, J. (2000a). Crystal structure of a D-aminopeptidase from *Ochrobactrum anthropi*, a new member of the 'penicillin-recognizing enzyme' family. Structure *8*, 971-980.

Bompard-Gilles, C., Villeret, V., Davies, G. J., Fanuel, L., Joris, B., Frère, J.-M., and van Beeumen, J. (2000b). A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-Ala-esterase/amidase from *Ochrobactrum anthropi*. Structure *8*, 153-162.

Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C. E., Smith, J. L., Tomchick, D. R., and Murzin, A. G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. Nature *378*, 416-419.

Bürgi, H. B., and Dunitz, J. D. (1983). From crystal statics to chemical dynamics. Acc. Chem. Res. *16*, 153-161.

Bürgi, H. B., Dunitz, J. D., Lehn, J. M., and Wipff, G. (1974). Stereochemistry of reaction paths at carbonyl centers. Tetrahedron *30*, 1563-1572.

Bush, K., and Jacoby, G. A. (2010). Updated functional classification of β-lactamases. Antimicrob. Agents Chemother. *54*, 969-976.

Chandel, A. K., Rao, L. V., Narasu, M. L., and Singh, O. V. (2008). The realm of penicillin G acylase in β-lactam antibiotics. Enzyme Microb. Technol. *42*, 199-207.

Cheng, H., and Grishin, N. V. (2005). DOM-fold: a structure with crossing loops found in DmpA, ornithine acetyltransferase, and molybdenum cofactor-binding domain. Protein Sci. *14*, 1902-1910.

Dawson, R. M. (1998). The toxicology of microcystins. Toxicon *36*, 953-962.

De Lano, W. L. (2002). The Pymol molecular graphics system (Palo Alto, CA, USA: DeLano Scientific).

Delmarcelle, M., Boursoit, M.-C., Filée, P., Baurin, S. L., Frère, J.-M., and Joris, B. (2005). Specificity inversion of *Ochrobactrum anthropi* D-aminopeptidase to a D,D-carboxypeptidase with new penicillin binding activity by directed mutagenesis. Protein Sci. *14*, 2296-2303.

Deshpande, A. D., Baheti, K. G., and Chatterjee, N. R. (2004). Degradation of β-lactam antibiotics. Curr. Sci. *87*, 1684-1695.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr., Sect. D: Biol. Crystallogr. *60*, 2126-2132.

Fanuel, L., Goffin, C., Cheggour, A., Devreese, B., van Driessche, G., Joris, B., van Beeumen, J., and Frère, J.-M. (1999). The DmpA aminopeptidase from *Ochrobactrum anthropi* LMG7991 is the prototype of a new terminal nucleophile hydrolase family. Biochem. J. *341*, 147-155.

Frackenpohl, J., Arvidsson, P. I., Schreiber, J. V., and Seebach, D. (2001). The outstanding biological stability of β- and γ-peptides toward proteolytic enzymes: an *in vitro* investigation with fifteen peptidases. ChemBioChem *2*, 445-455.

Fuchs, V., Jaeger, K.-E., Wilhelm, S., and Rosenau, F. (2011). The BapF protein from *Pseudomoas aeruginosa* is a β-peptidyl aminopeptidase. World J. Microbiol. Biotechnol. *27*, 713-718.

Geueke, B., Heck, T., Limbach, M., Nesatyy, V., Seebach, D., and Kohler, H.-P. E. (2006). Bacterial β-peptidyl aminopeptidases with unique substrate specificities for β-oligopeptides and mixed β,α-oligopeptides. FEBS J. *273*, 5261-5272.

Geueke, B., and Kohler, H.-P. E. (2007). Bacterial β-peptidyl aminopeptidases: on the hydrolytic degradation of β-peptides. Appl. Microbiol. Biotechnol. *74*, 1197-1204.

Geueke, B., Namoto, K., Seebach, D., and Kohler, H.-P. E. (2005). A novel β-peptidyl aminopeptidase (BapA) from strain 3-2W4 cleaves peptide bonds of synthetic β-tri- and β-dipeptides. J. Bacteriol. *187*, 5910-5917.

Guiotto, A., Calderan, A., Ruzza, P., and Borin, G. (2005). Carnosine and carnosine-related antioxidants: a review. Curr. Med. Chem. *12*, 2293-2315.

Heck, T., Kohler, H.-P. E., Limbach, M., Flögel, O., Seebach, D., and Geueke, B. (2007). Enzyme-catalyzed formation of β-peptides: β-peptidyl aminopeptidases BapA and DmpA acting as β-peptide-synthesizing enzymes. Chem. Biodiversity *4*, 2016-2030.

Heck, T., Limbach, M., Geueke, B., Zacharias, M., Gardiner, J., Kohler, H.-P. E., and Seebach, D. (2006). Enzymatic degradation of β- and mixed α,β-oligopeptides. Chem. Biodiversity *3*, 1325-1348.

Heck, T., Makam, V. S., Lutz, J., Blank, L. M., Schmid, A., Seebach, D., Kohler, H.-P. E., and Geueke, B. (2010a). Kinetic analysis of L-carnosine formation by β-aminopeptidases. Adv. Synth. Catal. *352*, 407-415.

Heck, T., Reimer, A., Seebach, D., Gardiner, J., Deniau, G., Lukaszuk, A., Kohler, H.-P. E., and Geueke, B. (2010b). β-Aminopeptidase-catalyzed biotransformations of β^2-dipeptides: kinetic resolution and enzymatic coupling. ChemBioChem *11*, 1129-1136.

Heck, T., Seebach, D., Osswald, S., ter Wiel, M. K. J., Kohler, H.-P. E., and Geueke, B. (2009). Kinetic resolution of aliphatic β-amino acid amides by β-aminopeptidases. ChemBioChem *10*, 1558-1561.

Juaristi, E. (2005). Enantioselective synthesis of β-amino acids (Hoboken: John Wiley & Sons).

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795-800.

Kiener, P. A., and Waley, S. G. (1978). Reversible inhibitors of penicillinases. Biochem. J. *169*, 197-204.

Komeda, H., and Asano, Y. (2005). A DmpA-homologous protein from *Pseudomonas* sp. is a dipeptidase specific for β-alanyl dipeptides. FEBS J. *272*, 3075-3084.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr., Sect. D: Biol. Crystallogr. *60*, 2256-2268.

Liljeblad, A., and Kanerva, L. T. (2006). Biocatalysis as a profound tool in the preparation of highly enantiopure β-amino acids. Tetrahedron *62*, 5831-5854.

Luukkainen, R., Sivonen, K., Namikoshi, M., Färdig, M., Rinehart, K. L., and Niemelä, S. I. (1993). Isolation and identification of 8 microcystins from thirteen *Oscillatoria agardhii* strains and structure of a new microcystin. Appl. Environ. Microbiol. *59*, 2204-2209.

Matagne, A., Dubus, A., Galleni, M., and Frère, J.-M. (1999). The β-lactamase cycle: a tale of selective pressure and bacterial ingenuity. Nat. Prod. Rep. *16*, 1-19.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658-674.

Oinonen, C., and Rouvinen, J. (2000). Structural comparison of Ntn-hydrolases. Protein Sci. *9*, 2329-2337.

Powers, J. C., Asgian, J. L., Ekici, Ö. D., and James, K. E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. Chem. Rev. *102*, 4639-4750.

Seebach, D., Beck, A. K., and Bierbaum, D. J. (2004). The world of β- and γ-peptides comprised of homologated proteinogenic amino acids and other components. Chem. Biodiversity *1*, 1111-1239.

Seebach, D., and Gardiner, J. (2008). β-Peptidic peptidomimetics. Acc. Chem. Res. *41*, 1366-1375.

Steer, D. L., Lew, R. A., Perlmutter, P., Smith, A. I., and Aguilar, M.-I. (2002). β-Amino acids: versatile peptidomimetics. Curr. Med. Chem. *9*, 811-822.

Suda, H., Takita, T., Aoyagi, T., and Umezawa, H. (1976). Structure of bestatin. J. Antibiot. *29*, 100-101.

Umezawa, H., Aoyagi, T., Suda, H., Hamada, M., and Takeuchi, T. (1976). Bestatin, an inhibitor of aminopeptidase-B, produced by actinomycetes. J. Antibiot. *29*, 97-99.

Wiegand, H., Wirz, B., Schweitzer, A., Camenisch, G. P., Rodriguez Perez, M. I., Gross, G., Woessner, R., Voges, R., Arvidsson, P. I., Frackenpohl, J., and Seebach, D. (2002). The outstanding metabolic stability of a ^{14}C-labeled β-nonapeptide in rats – *in vitro* and *in vivo* pharmacokinetic studies. Biopharm. Drug Dispos. *23*, 251-262.

FIGURE LEGENDS

Figure 1: Similarities of L-α-amino acids and β-amino acids of corresponding configurations

According to the nomenclature proposed by Seebach and coworkers β-amino acids with proteinogenic side chains (R) are referred to as β-homoamino acids (βhXaa residues) (Seebach et al., 2004). As with α-amino acids (Xaa residues), the notations H-β^3hXaa-OH and H-β^2hXaa-OH implicitly refer to the configurations of the stereocenters as shown in the formulae, *i.e.* L in the *Fischer* projection (strictly applied, the rules for this projection would lead to D-configuration only in the case of β^2hAla). The CIP nomenclature (*R*/*S*-configuration) is not useful for amino acids in a biological context because it reverses, depending on the structures of the side chains. Thus, all Xaa residues have (*S*)-configuration, except for Cys and Sec; β^3hXaa residues have (*S*)-configuration, except for Val, Ile, Ser, Thr, Cys, Sec, Met and Asp; β^2hXaa residues have (*R*)-configuration, except for Ser and Thr.

Figure 2: Chemical formulae of AEBSF and β-lactam-derived compounds

Hydrolysis of the β-lactam amide bond of ampicillin by a β-lactamase leads to formation of the linear penicilloic acid (5*R*)-Amp_{hyd}, which subsequently undergoes epimerization at C-5 of the thiazolidine ring to form (5*S*)-Amp_{hyd}, the prevailing epimer in aqueous solution (as shown by NMR, see Supplemental Information available online). 6-APA is the common reaction product formed by the penicillin-acylase-catalyzed hydrolysis of penicillin-derived β-lactam antibiotics. The β-amino-acid moieties of the compounds are highlighted in red.

Figure 3: Active site of BapA in complex with the serine-protease inhibitor AEBSF

The essential catalytic residues of BapA (subunit A) and residues contributing to the hydrophobic ligand-binding pocket are shown as sticks in yellow and grey, respectively. The subunit localizations of the binding-pocket residues are indicated with subscript letters. AEBSF is depicted as sticks in light blue and the catalytically relevant water molecule is shown as a red sphere. The catalytic residues and AEBSF are contoured with a 2 Fofc electron-density map at 1.5 σ. Interactions between the amino group of AEBSF and the active site of BapA are indicated by dotted lines, distances are given in Å.

Figure 4: Active site of BapA in complex with β-lactam-derived inhibitors

(A) BapA in complex with ampicillin.

(B) BapA in complex with the pencilloic acid Amp_{hyd}.

The catalytic residues of BapA (subunit A) and selected residues of the ligand-binding pockets are shown as sticks in yellow and grey, respectively. The subunit localizations of the binding-pocket residues are indicated with subscript letters. Ampicillin and Amp_{hyd} are depicted as sticks in light blue and dark blue, respectively, contoured with a 2 Fofc electron-density map at 1.5 σ. The catalytically relevant water molecule is shown as a red sphere. Interactions between the bound compounds and active-site residues of BapA are indicated by dotted lines, distances are given in Å.

Figure 5: Models of tetrahedral BapA-substrate intermediates

(A) Tetrahedral reaction intermediates between Ser250-Oγ and β^3-homophenylalanine residues of L-configuration ((+)-sc conformer; orange) and D-configuration ((-)-sc conformer; magenta). The models are based on the BapA-Amp_{hyd} crystal structure (grey); the ligand Amp_{hyd} is indicated as lines contoured with a 2 Fofc electron-density map at 1.5 σ. The catalytic residues of BapA are shown as yellow sticks and the catalytically relevant water molecule as a red sphere. The predicted salt bridge and hydrogen bonds of the oxyanion hole are indicated by dotted lines.

(B) Superposition of the active sites of BapA-Amp_{hyd} (grey) on the crystal structure of DmpA (PDB ID: 1b65; dark red) shown in ribbon presentation. The catalytic residues Glu133, Ser250, Ser288 and Glu290 of BapA (yellow sticks) correspond to Glu144, Ser250, Ser288 and Asp290 of DmpA (red sticks). The side chain of Trp137 of DmpA clashes with the BapA ligands AEBSF (light blue), Amp_{hyd} (dark blue), and the modeled BapA-L-β^3-homophenylalanine intermediate (orange; see **A**).

Tetrahedral reaction intermediates between BapA Ser250-Oγ and the (+)-sc conformers of D- and L-β^2-homophenylalanine residues are presented in **Figure S2** (available online).

TABLE LEGENDS

Table 1: Data collection and refinement statistics (molecular replacement)

Values in parentheses represent data for the highest resolution shell. One crystal was used for each dataset. [a] $R_{sym} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I is the measured intensity of each reflection and <I> is the intensity from multiple observations of symmetry-related reflections. [b] R-factor $= \sum |F_{obs}| - |F_{calc}| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are the observed and the calculated structure factor amplitudes, respectively. [c] R_{free} is the R-factor calculated with reflections chosen at random and omitted from refinement.

Table 2: Inhibition of the BapA by the β-lactam derivatives ampicillin and Amp_{hyd}

The kinetic parameters K_m, V_{max} and K_i of the reactions were obtained after fitting initial velocities of the b-aminopeptidase-catalyzed conversion of the chromogenic substrate H-β^3hAla-*p*NA (pH 7.2, 37°C) (Heck et al., 2006) to a competitive inhibition model (Eq. 1, Experimental Procedures). Penicillin G and carbenicillin did not inhibit substrate conversion by BapA.

Table 3: Survey of the conformers of BapA-substrate intermediates

The tetrahedral reaction intermediates are formed through nucleophilic attack of Ser250-Oγ of BapA to β^3-homophenylalanine residues of L- or D-configuration (see also **Figure S1**).

TABLES

Table 1: Data collection and refinement statistics (molecular replacement)

	BapA-AEBSF	BapA-Amp	BapA-Amp$_{hyd}$
Data collection			
Space group	P21	P21	P21
Cell parameters: *a*, *b*, *c* (Å)	87.4, 96.8, 101.3, β = 108.4°	86.9, 96.4, 101.5, β = 108.4°	88.3, 97.1, 102.2, β = 108.7°
Resolution range (Å)	48.0 – 1.8 (1.9 – 1.8)	48.8 – 1.7 (1.8 – 1.7)	49.3 – 1.85 (1.9 – 1.85)
Completeness (%)	97.6 (94.7)	98.7 (94.4)	98.4 (99.2)
Redundancy	3.78 (3.70)	3.32 (2.73)	3.78 (3.77)
R_{sym} [a]	10.0 (49.7)	11.6 (58.1)	8.6 (45.5)
$I / \sigma I$	12.9 (3.1)	9.2 (2.2)	13.0 (4.0)
Refinement			
Resolution (Å)	1.8	1.7	1.85
N° unique reflections	144691	172065	137077
N° free reflections	1448	6884	1371
R_{work} [b] / R_{free} [c]	16.0 / 19.0	16.7 / 19.2	15.0 / 17.8
N° of atoms			
Protein	10708	10668	10680
Ligand / ions	238	168	194
Water	1265	1138	1229
B-factors			
Protein	19.61	18.81	19.58
Ligand/ion	27.85	25.61	26.17
Water	29.48	26.53	29.50
R.m.s. deviations			
Bond lengths (Å)	0.007	0.008	0.007
Bond angles (°)	1.05	1.104	1.05
Ramachandran			
Most favored	1391 (96.7%)	1385 (96.6)	1375 (96.4%)
Generously allowed	43 (3.0%)	44 (3.1%)	48 (3.3%)
Disallowed	4 (0.3%)	5 (0.3%)	4 (0.3%)

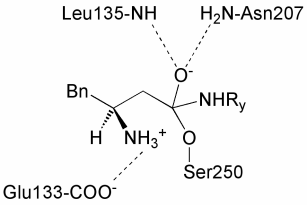
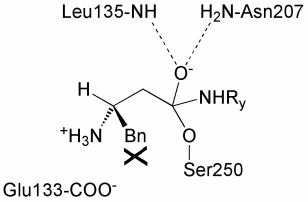
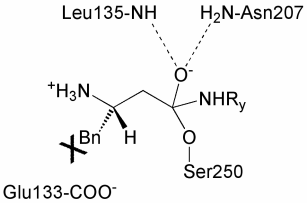
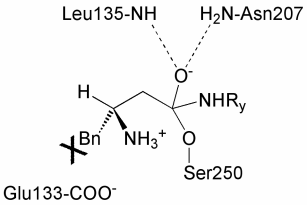
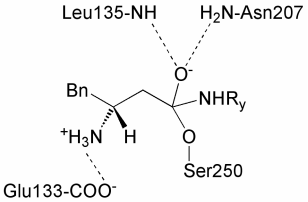
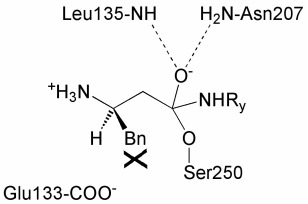
Values in parentheses represent data for the highest resolution shell. One crystal was used for each dataset. [a] $R_{sym} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ where I is the measured intensity of each reflection and $\langle I \rangle$ is the intensity from multiple observations of symmetry-related reflections. [b] R-factor $= \sum |F_{obs}| - |F_{calc}| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are the observed and the calculated structure factor amplitudes, respectively. [c] R_{free} is the R-factor calculated with reflections chosen at random and omitted from refinement.

Table 2: Inhibition of the BapA by the β-lactam derivatives ampicillin and Amp_{hyd}

Inhibitor	K_i [mM]	K_m [mM]	V_{max} [μmol·$min^{-1}$$mg^{-1}$]
Without	-	1.1 ± 0.15	16.4 ± 0.86
Ampicillin	0.69 ± 0.05	1.1 ± 0.09	16.3 ± 0.50
Amp_{hyd}	0.74 ± 0.11	0.98 ± 0.16	16.8 ± 1.0

The kinetic parameters K_m, V_{max} and K_i of the reactions were obtained after fitting initial velocities of the β-aminopeptidase-catalyzed conversion of the chromogenic substrate H-β^3hAla-*p*NA (pH 7.2, 37°C) (Heck et al., 2006) to a competitive inhibition model (Eq. 1, Experimental Procedures). Penicillin G and carbenicillin did not inhibit substrate conversion by BapA.

Table 3: Survey of the conformers of BapA-substrate intermediates

Conformation[a]		Color code in **Figure S1**	Clashes between Cγ / Cδ and BapA residues	Salt bridge to Glu133
L-β^3-Homophenyalanine residue				
(+)-sc		Orange	Cγ: - Cδ: -	+
(-)-sc		Magenta	Cγ: - Cδ: Thr76, Gly289	-
ap		Green	Cγ: Leu135 Cδ: Leu135, Thr100, Glu133	-
D-β^3-Homophenyalanine residue				
(+)-sc		Orange	Cγ: Glu133 Cδ: Glu133, Thr134, Leu135	-
(-)-sc		Magenta	Cγ: - Cδ: -	+
ap		Green	Cγ: Thr76, Gly289 Cδ: Thr76, Gly289	-

The tetrahedral reaction intermediates are formed through nucleophilic attack of Ser250-Oγ of BapA to β^3-homophenylalanine residues of L- or D-configuration (see also **Figure S1**).

[a] Dotted lines represent non-covalent interactions. **X** indicates steric repulsions between the side chain of the substrate and residues of the BapA active site. R_y represents the C-terminal part of the substrate prior to formation of the acyl enzyme (see Merz et al., this issue).

FIGURES

Figure 1

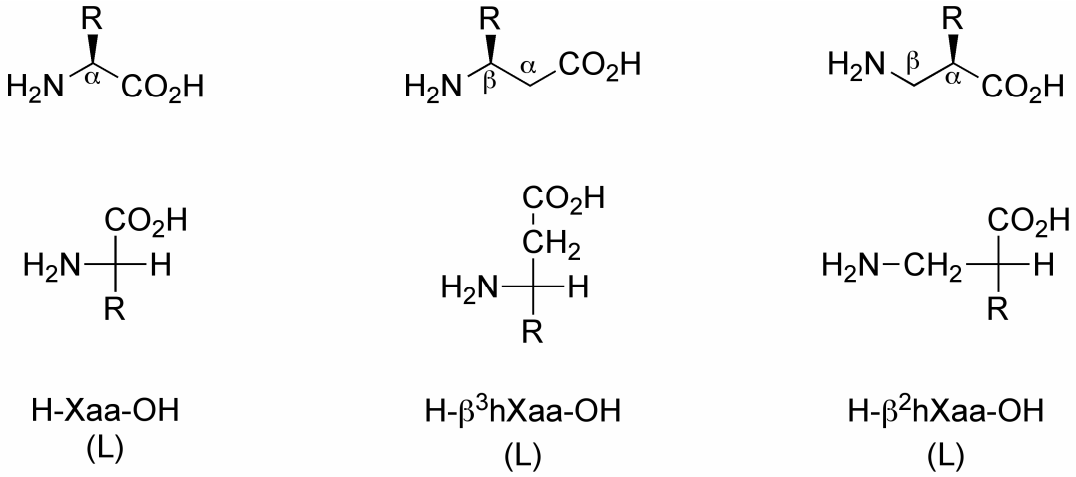


Figure 2

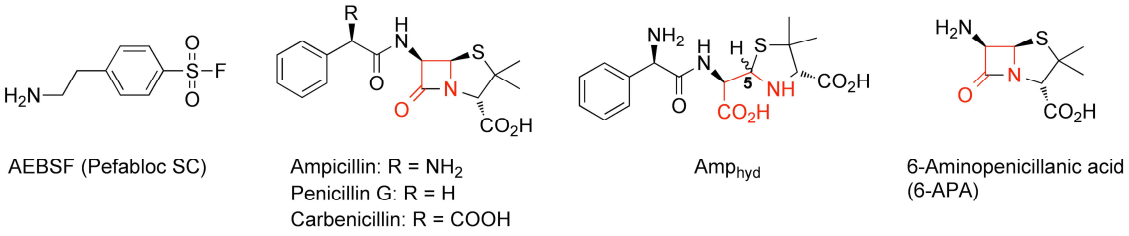


Figure 3

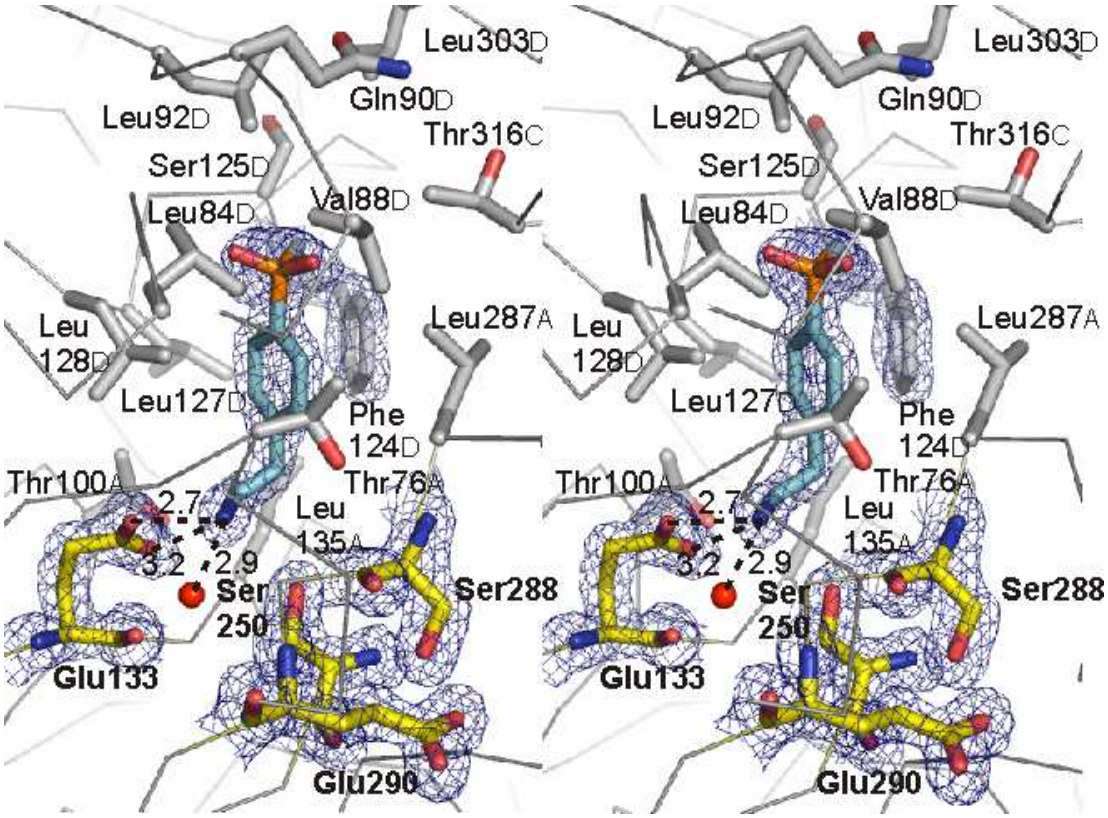

Leu303D
Gln90D
Leu92D
Thr316C
Ser125D
Leu84D
Val88D
Leu
128D
Leu287A
Leu127D
Phe
124D
Thr100A
Thr76A
2.7
Leu
135A
3.2
2.9
Ser
250
Ser288
Glu133
Glu290

Figure 4A

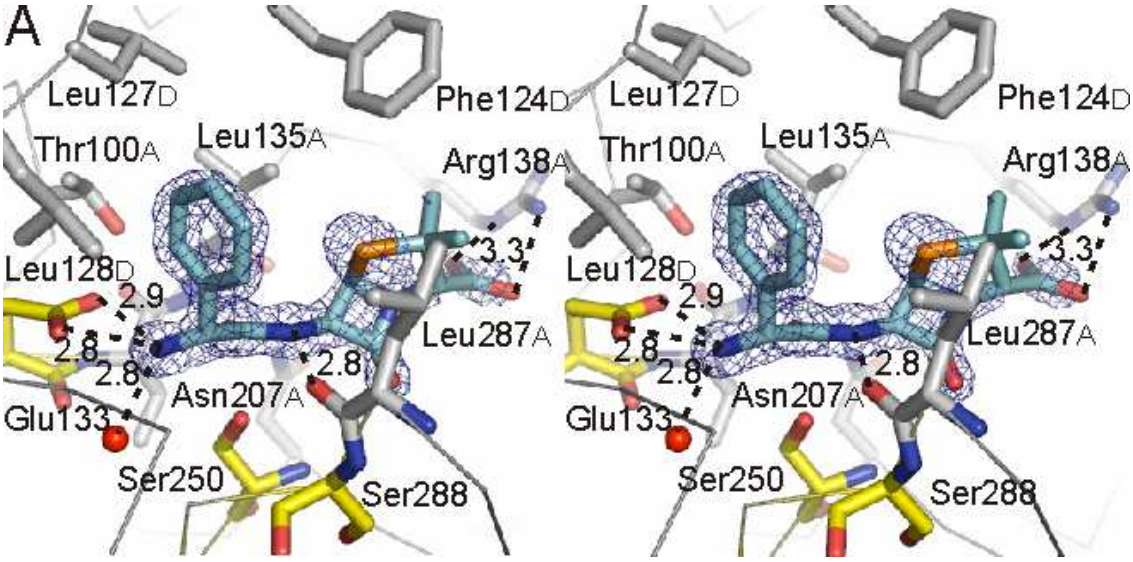


Figure 4B

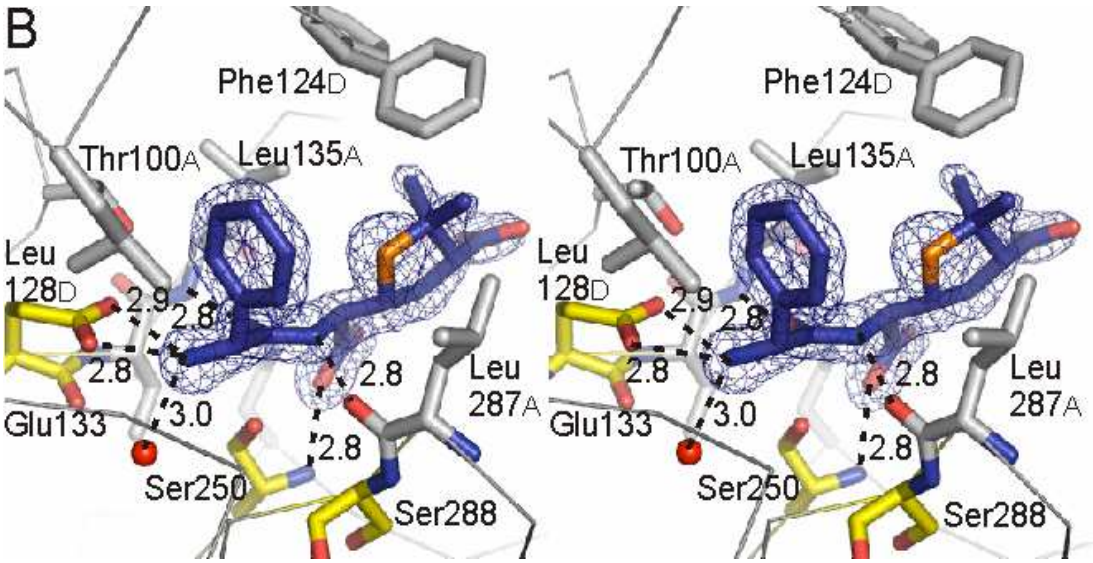


Figure 5A

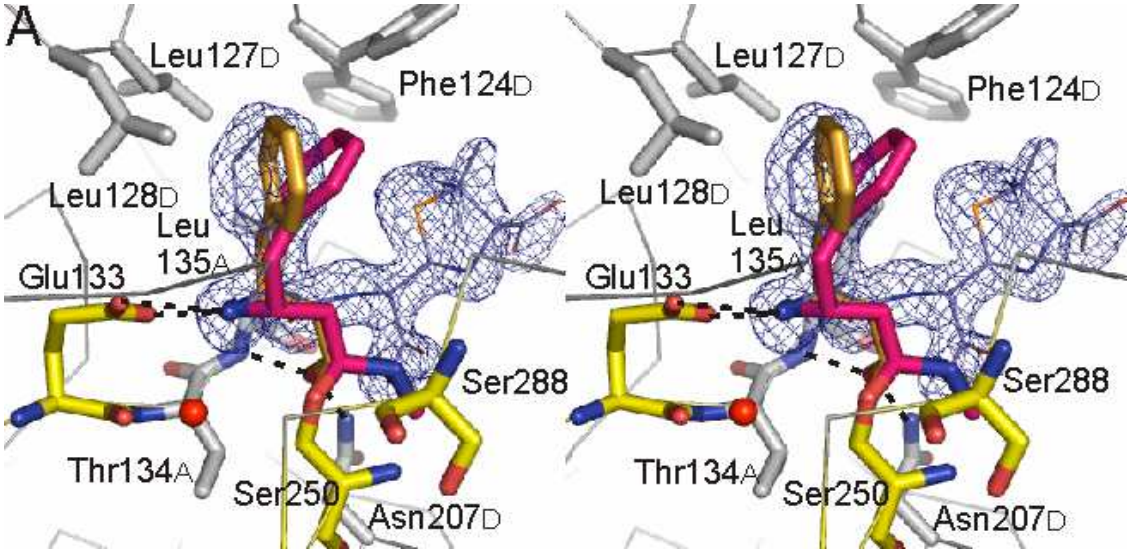
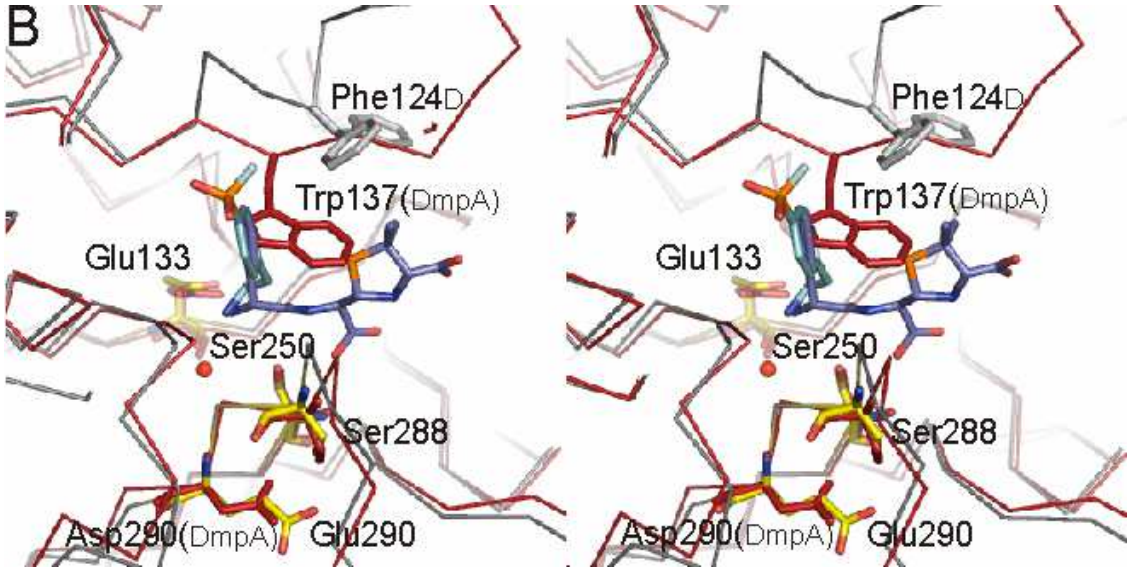


Figure 5B



SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures showing models of tetrahedral reaction intermediates of BapA Ser250-Oγ with β^3-homophenylalanine residues (Figure S1) and β^2-homophenylalanine residues (Figure S2). The NMR analysis of the β-lactam derivatives used in this study is presented in the Supplemental Experimental Procedures.

Supplemental Figures

Figure S1: Models of tetrahedral reaction intermediates formed between BapA Ser250-Oγ and the energetically favorable conformations of L- and D-β^3-homophenylalanine residues, related to Table 3

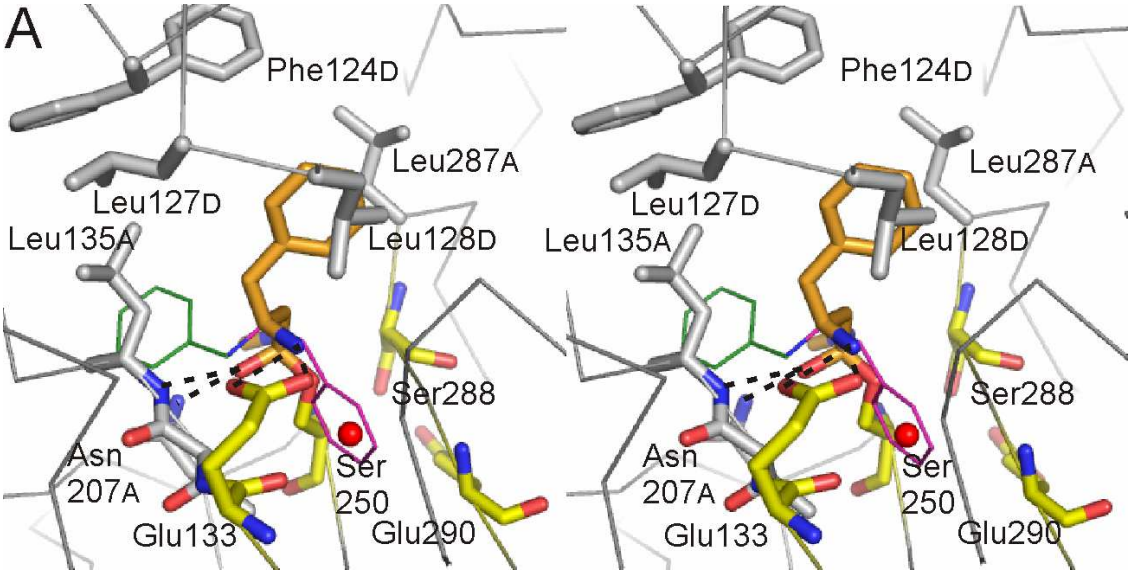


Figure S1B

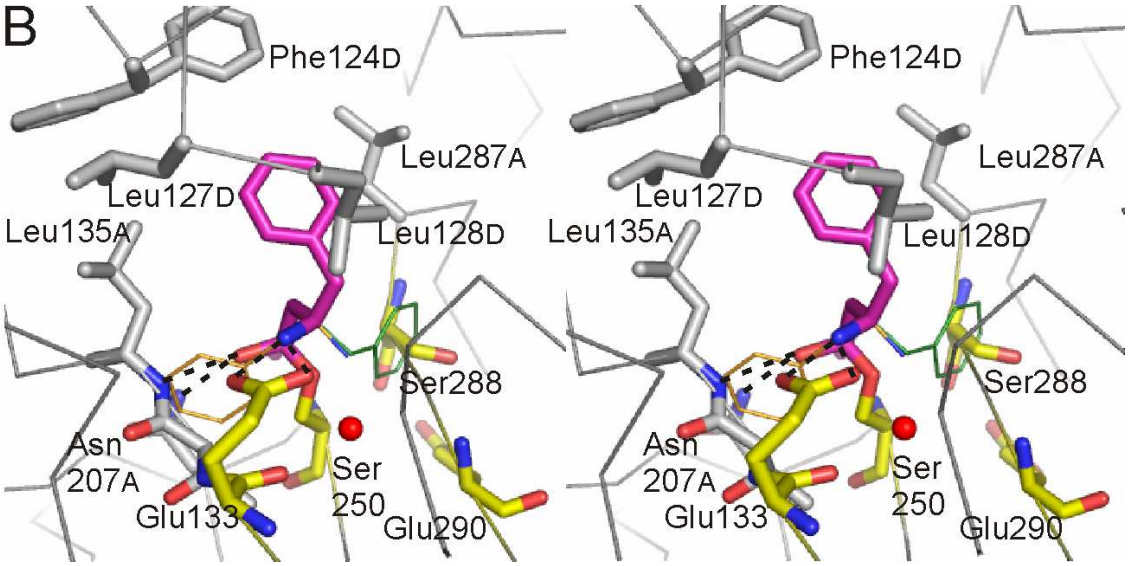


Figure S1: **Models of BapA-substrate intermediates**

(A) Tetrahedral reaction intermediates formed between BapA Ser250-Oγ and L-β^3-homophenylalanine.

(B) Tetrahedral reaction intermediates formed between BapA Ser250-Oγ and D-β^3-homophenylalanine.

The models are based on the crystal structure of BapA-Amp_{hyd} (grey). The catalytic residues of BapA (subunit A) and selected residues of the ligand-binding pocket are highlighted as sticks in yellow and grey, respectively; the subunit localizations of the labeled residues are indicated with subscript letters. The (+)-sc, (-sc) and ap conformers of the ligands are shown in orange, magenta and green, respectively; the conformer that best matches the active site of BapA is highlighted as sticks. The predicted salt bridge and hydrogen bonds of the oxyanion hole are indicated by dotted lines. For a schematic overview of the depicted conformers we refer to **Table 3** in the main text.

Figure S2: Models of tetrahedral reaction intermediates formed between BapA Ser250-Oγ and the (+)-sc conformers of L- and D-β²-homophenylalanine residues, related to Figure 5

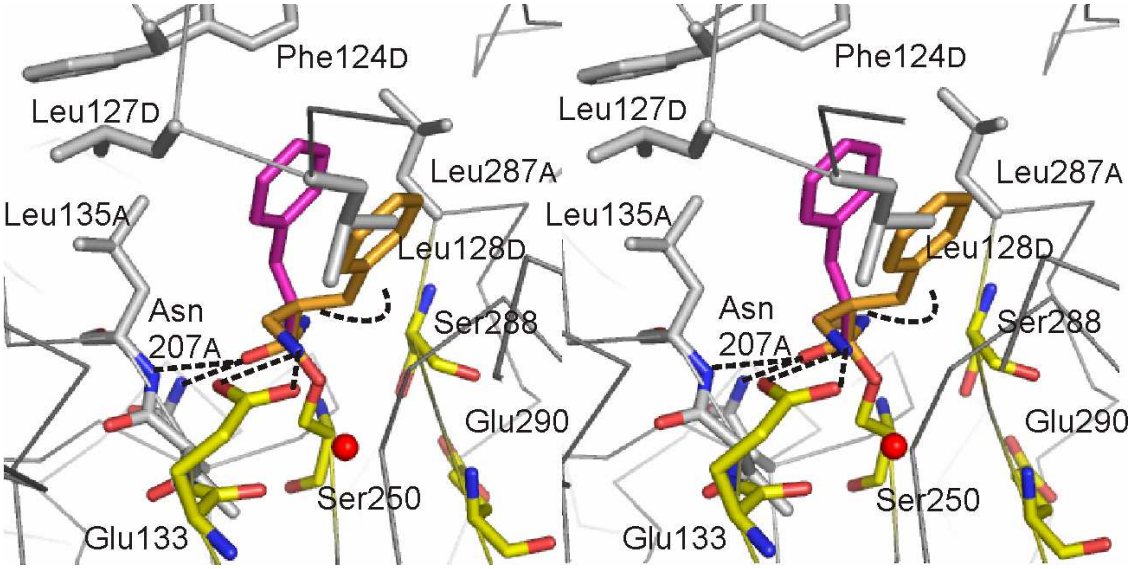


Figure S2: Models of tetrahedral reaction intermediates formed between BapA Ser250-Oγ and β²-homophenylalanine residues

The models are based on the crystal structure of BapA-Amp_{hyd} (grey). The (+)-sc conformers of the bound D- and L-β²-homophenylalanine residues are shown as sticks in magenta and orange, respectively. The catalytic residues of BapA (subunit A) and selected residues of the ligand-binding pocket are highlighted as sticks in yellow and grey, respectively; the subunit localizations of the labeled residues are indicated with subscript letters. The predicted salt bridge and hydrogen bonds of the oxyanion hole are indicated by straight, steric repulsions by bent dotted lines.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

NMR analysis of Amp_{hyd}

The ^{1}H, ^{13}C and ^{15}N NMR spectra used for stereochemical assignments were recorded at 400.13, 100.61 and 40.56 MHz on a Bruker Avance-400 NMR spectrometer (Bruker Biospin AG, Fällanden, Switzerland). The ^{1}H, the 1D NOESY spectra and the 1D ^{13}C spectra, and the ^{1}H ^{13}C HSQC, ^{1}H ^{13}C HMBC, ^{1}H ^{1}H DQF-COSY and ^{1}H ^{15}N HMQC (recorded with ^{15}N at natural abundance) 2D correlation NMR experiments were performed at 298 K on a 5 mm broadband inverse probe with z-gradient (100% gradient strength of 53.5 Gcm^{-1}) and 90° pulse lengths of 6.8 (^{1}H), 14.5 (^{13}C) and 21.3 ⌈s (^{15}N). All NMR experiments were performed in D_2O solutions using the Bruker standard parameter sets and pulse programs selecting coupling constants of 145 Hz (HSQC), 10 Hz (HMBC), 3 Hz (HMQC) and mixing times of 1 s (1D NOESY, applying 50 ms selective 180º inversion pulses). The ^{1}H and ^{13}C chemical shifts were referenced relative to the signals of 3-trimethylsilyl tetradeutero sodium propionate (TSP) dissolved in D_2O at 0.0 and -1.6 ppm, respectively. The ^{15}N chemical shifts were referenced to a capillary containing neat nitromethane at 0.0 ppm. The relative amounts of products were determined by integration of ^{1}H NMR signals of the H-9 / H-10 methyl groups.

The ^{1}H and ^{13}C chemical shifts of the starting material ampicillin were readily assigned over the 2D correlated NMR experiments (**Table S1** and **S2**), and the relative configuration of H-9 and H-10 determined by NOE enhancements found between the pairs of protons H-3 / H-10 and H-5 / H-9, whereas no close contacts were found for H-5 / H-10 (**Table S3**). The ^{1}H chemical shifts evaluated for H-9 and H-10 completely agree with an earlier study (Tung et al., 2000). The ^{15}N chemical shift of N-4 at -213 ppm is typical for the β-lactam nitrogen

(**Table S1**). For the structurally closely related penicillin G a value of -212 ppm was observed (Dodd et al., 2010). After some days in solution, additional ^{1}H NMR resonances belonging to diastereotopic methyl groups were observed. These signals were assigned to a primary ((5*R*)-Amp_{hyd}, amount in the 1% region) and a secondary hydration product ((5*S*)-Amp_{hyd}, traces). The resonances of these two products were observed as the main signals in the ^{1}H NMR spectrum, when ampicillin was converted enzymatically by a β-lactamase (see Experimental Procedures). Two additional sets of ^{1}H and ^{13}C NMR resonances similar to the data found for ampicillin were found for the hydration products (5*R*)- and (5*S*)-Amp_{hyd} (**Table S1**). Whereas all ^{13}C chemical shifts of the individual carbons and also the ^{15}N NMR data of N-11 and N-18 in these molecules remained very similar to the starting material ampicillin (**Table S1**), a low frequency shift of approximately 100 ppm was observed for N-4 of the hydration products compared with ampicillin, apparently induced by the ring opening reaction at of the β-lactam moiety. The relative configuration of the primary hydration product (5*R*)-Amp_{hyd} was established by NOE enhancements found between the pairs of protons H-3 / H-10 and H-5 / H-9, whereas no close contacts were found for H-5 / H-10 (the same NOE's have previously been observed for ampicillin; **Table S3**). The NOE contact observed between H-3 and H-6 implies a rotation around the C-5 / C-6 bond of the compound (**Figure S3**). For the second hydration product (5*S*)-Amp_{hyd} an NOE enhancement was observed between H-3 and H-5 (and also to H-10), indicating an epimerization at C-3 or C-5. In fact, this kind of reaction is already mentioned in the literature (Deshpande et al., 2004; Robinson Fuentes et al., 1997); the postulated reaction mechanism is shown in **Scheme S1**. The ratio of (5*R*)-Amp_{hyd} / (5*S*)-Amp_{hyd} shifted from 36 / 64 % (initial) to 15 / 85 % over 9 days. It has to be mentioned that further, not identified by-products were formed during this time period (amounts in the low % range).

SUPPLEMENTAL TABLES

Table S1: **^{1}H, ^{13}C and ^{15}N chemical shifts of ampicillin and its derivatives (5*R*)-Amp$_{hyd}$ and (5*S*)-Amp$_{hyd}$**

Position[a]	δ (^{1}H) / ppm			δ (^{13}C or ^{15}N) / ppm		
	Ampicillin	(5*R*)-Amp$_{hyd}$	(5*S*)-Amp$_{hyd}$	Ampicillin	(5*R*)-Amp$_{hyd}$	(5*S*)-Amp$_{hyd}$
2				65.3	59.7	59.6
3	4.19	3.09	3.32	74.3	76.4	76.7
5	5.48	5.07	5.04	67.7	66.7	67.6
6	5.47	4.29	4.89	59.0	60.9	56.4
7				175.6	176.8	175.6
8				175.6	176.4	175.5
9	1.44	1.14	0.52	27.5	27.2	28.2
10	1.52	1.21	1.48	31.3	27.0	29.0
12				176.4	169.9	170.9
13	4.68	5.21	5.28	59.2	58.1	58.3
14				140.1	133.5	133.5
15	7.43	7.58	7.58	128.1	129.2	129.1
16	7.45	7.55	7.55	130.2	130.8	131.3
17	7.40	7.52	7.52	129.6	131.5	131.5
4				-213.2	-322.6	-310.6
11				-269.8	-258.4	-264.0
18				-349.8	n.d.[b]	-339.4

[a] For numbering of carbon atoms see **Figure S3**.

[b] No correlation observed (signal too weak or ^{1}H,^{15}N coupling constant too small for efficient polarization transfer).

Table S2: ^{1}H ^{13}C HMBC and ^{1}H ^{15}N HMQC correlations used for NMR shift assignment in Table S1

Compound	Observed HMBC and HMQC correlations (w: weak)
Ampicillin	H-3→C-(2, 5, 8, 9w, 10); H-5→C-(6, 7); H-6→C-(5, 7, 12w); H-9→C-(2, 3, 9); H-10→C-(2, 3, 10); H-13→C-(12, 14, 15); H-15→C-(13, 14, 15, 17); H-16→C-(14, 16); H-17→C-(15) H-(3, 5, 6)→N-(4); H-13→N-(11, 18)
(5*R*)-Amp$_{hyd}$	H-3→C-(2, 8, 9, 10w); H-5→C-(2w, 3, 7); H-6→C-(5, 7, 12w); H-9→C-(2, 3, 10); H-10→C-(2, 3, 9); H-13→C-(12, 14, 15); H-16→C-(14, 16); H-17→C-(15) H-3→N-(4); H-5→N-(4); H-6→N-(4, 11)
(5*S*)-Amp$_{hyd}$	H-3→C-(2, 8, 9, 10w); H-5→C-(7w); H-6→C-(5, 7, 12w); H-9→C-(2, 3, 10); H-10→C-(2, 3, 9); H-13→C-(12, 14, 15); H-15→C-(13, 15, 17); H-16→C-(14, 16); H-17→C-(15) H-3→N-(4); H-(5, 6)→N-(4, 11); H-13→N-(18)

Table S3: ^{1}H 1D NOESY signals used for assignments of stereochemistry

Compound	Observed 1D-NOESY correlations (w: < 1% NOE observed)
Ampicillin	H-3→H-(10); H-9→ H-(3, 5); H-10→ H-(3, 9)
(5*R*)-Amp$_{hyd}$	H-3→H-(6, 10, 15); H-5→H-(3w, 6, 9w); H-6→H-(3, 5); H-9→H-(3, 5, 10); H-10→H-(3, 9); H-13→H-(15)
(5*S*)-Amp$_{hyd}$	H-3→H-(5, 10); H-5→H-(3, 6); H-6→H-(5, 13w); H-9→H-(3, 10, 15); H-10→H-(3, 9); H-13→H-(6w, 15)

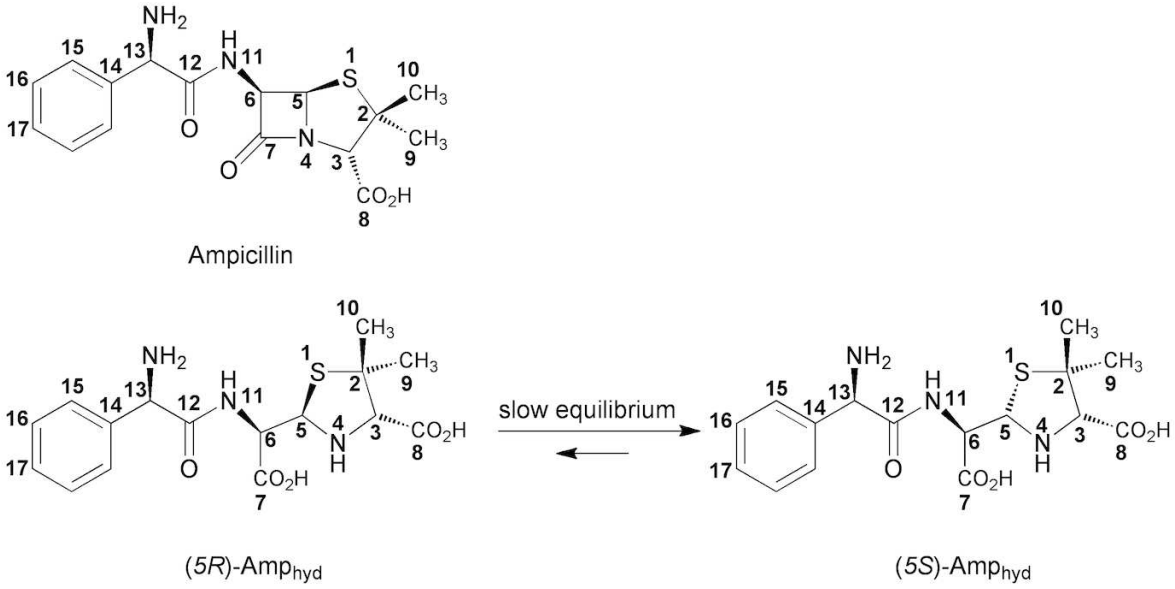
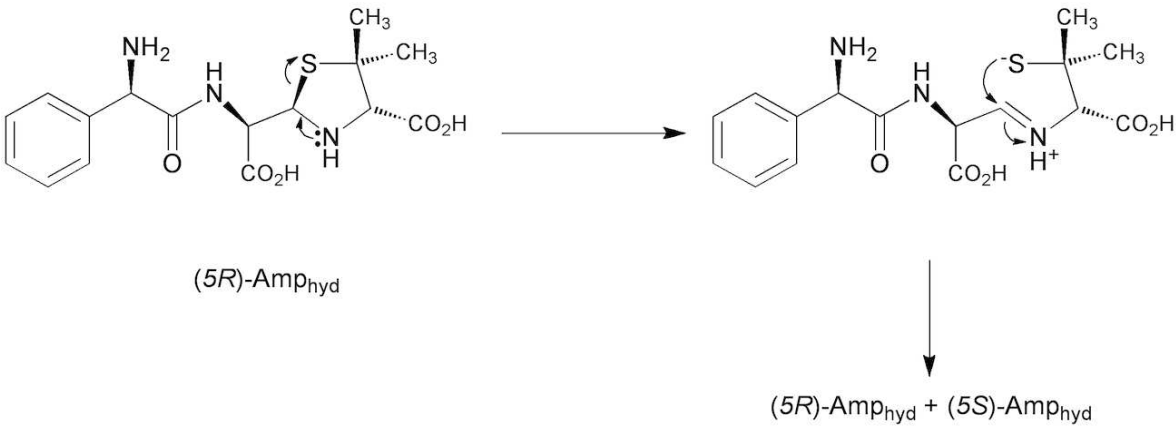


Figure S3: Numbering of carbon atoms used for the NMR signal assignment of ampicillin and its derivatives (5*R*)-Amp$_{hyd}$ and (5*S*)-Amp$_{hyd}$ (as employed by Kiener and Waley, 1978)



Scheme S1: Proposed reaction mechanism of the epimerization at C-5 of Amp$_{hyd}$

SUPPLEMENTAL REFERENCES

Dodd, M.C., Rentsch, D., Singer, H.P., Kohler, H.-P.E., and von Gunten, U. (2010). Transformation of β-lactam antibacterial agents during aqueous ozonation: Reaction pathways and quantitative bioassay of biologically-active oxidation products. Environ. Sci. Technol. *44*, 5040-5049.

Robinson Fuentes, V.A., Jefferies, T.M., and Branch, S.K. (1997). Degradation pathways of ampicillin in alkaline solutions. J. Pharm. Pharmacol. *49*, 843-851.

Tung, J.C., Gonzales, A.J., Sadowsky, J.D., and O'Leary, D.J. (2000). On the ^{1}H NMR chemical shift assignments of ampicillin, Magn. Reson. Chem. *38*, 126-128.

CHAPTER 4

NMR STUDY OF A HOMOTYPIC NLRP1-PYD - ASC-PYD INTERACTION

INTRODUCTION

The NLRP1- and NLRP3 inflammasomes[1] the apoptosome, the NOD-signalosome[2] and the PIDDosome[3] are intracellular pathogen sensing complexes. The membrane bound DISC[4], TRAIL[5] and Toll-like receptors are complexes that sense extracellularly occurring bacterially derived components [1], [2], [3], [4]. All these molecules share several common structural features. They possess a microbial ligand-sensing domain, which is often made up of a series of leucine rich repeats, a protein-protein interaction domain that is used for the recruitment and activation of signal transducers such as caspases and kinases [5], [6] as well as relatively small domains, called death domains, which are postulated to be involved in homotypic interactions with partner molecules. They represent the "glue" of these multimeric protein complexes. According to their similar functional characteristics and the structural domain arrangement, the molecules are combined into the group of pathogen associated molecular pattern[6] recognition molecules [4]. The detection of bacterially derived components in the extracellular or intracellular environment stimulates these PAMP molecules to assemble into large multimeric complexes, often termed activation platforms. They are responsible for the regulation of downstream signalling processes. The detection of microorganisms is highly efficient because structural components like lipopolysaccharides, muramyl peptides and other bacterial cell wall components remained conserved during evolution [7].

Molecules that recognize pathogen associated patterns are rather large and flexible and possess unfavourable biophysical properties to be structurally investigated. Consequently, the focus in structural research is predominantly based on the functional units, which comprise the four different types of death domains as well as the caspase entity, the helicase entity, the ligand sensing leucine rich repeat and the oligomerization inducing nucleotide binding domain, which also exhibits essential inflammatory and apoptotic functions. This chapter focuses on the pyrin[7] death domain interactions of NLRP1 and the adaptor protein apoptosis associated speck-like protein containing a CARD[8], which both are components of the multi-protein complex termed inflammasome.

[1] Nucleotide binding and leucine rich repeat containing protein; NLRP
[2] Nucleotide binding and oligomerization domain; NOD
[3] P53 induced protein with a death domain; PIDD
[4] Death inducing signalling complex; DISC
[5] Tumour necrosis factor-related apoposis-inducing ligand receptor; TRAIL
[6] Pathogen associated molecular pattern; PAMP
[7] Pyrin; PYD
[8] Apoptosis associated speck-like protein containing a CARD; ASC

1. Superfamily of death domain fold[9] containing proteins

The superfamily of death domain fold containing proteins comprises four families according to the SCOP database:

DD: death domain family
DED: death effector domain family
CARD: caspase activation and recruitment domain family
PYD: pyrin family

The common motif of all four DDF families is an antiparallel six helix bundle of amphipathic helices with an approximate size of 85-100 amino acids (Figure 1_4).

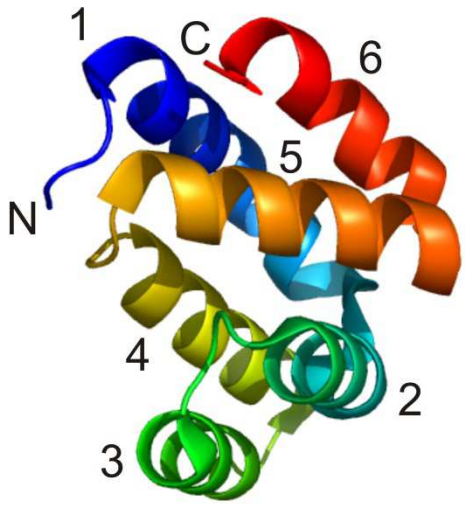


Figure 1_4)
Cartoon representation of a representative DDF structure (the CARD domain of Apaf-1; protein database code 2P1H). The helices are numbered from 1 to 6, *N* and *C* denote the amino- and carboxy-termini, respectively.

The individual lengths and orientations of the six α-helices vary within and between each family. The sequence identity of all proteins among the four families is low (between 10% and 25%). Due to relatively high diversity of the domains, experimental structure determination as well as protein-protein interaction studies of different members of the families are a prerequisite for a detailed understanding of the structural and functional relationships of these domains.

A selection of the most important DDF containing proteins, used in the following sections, is shown in Figure 2_4).

[9] Death domain fold; DDF

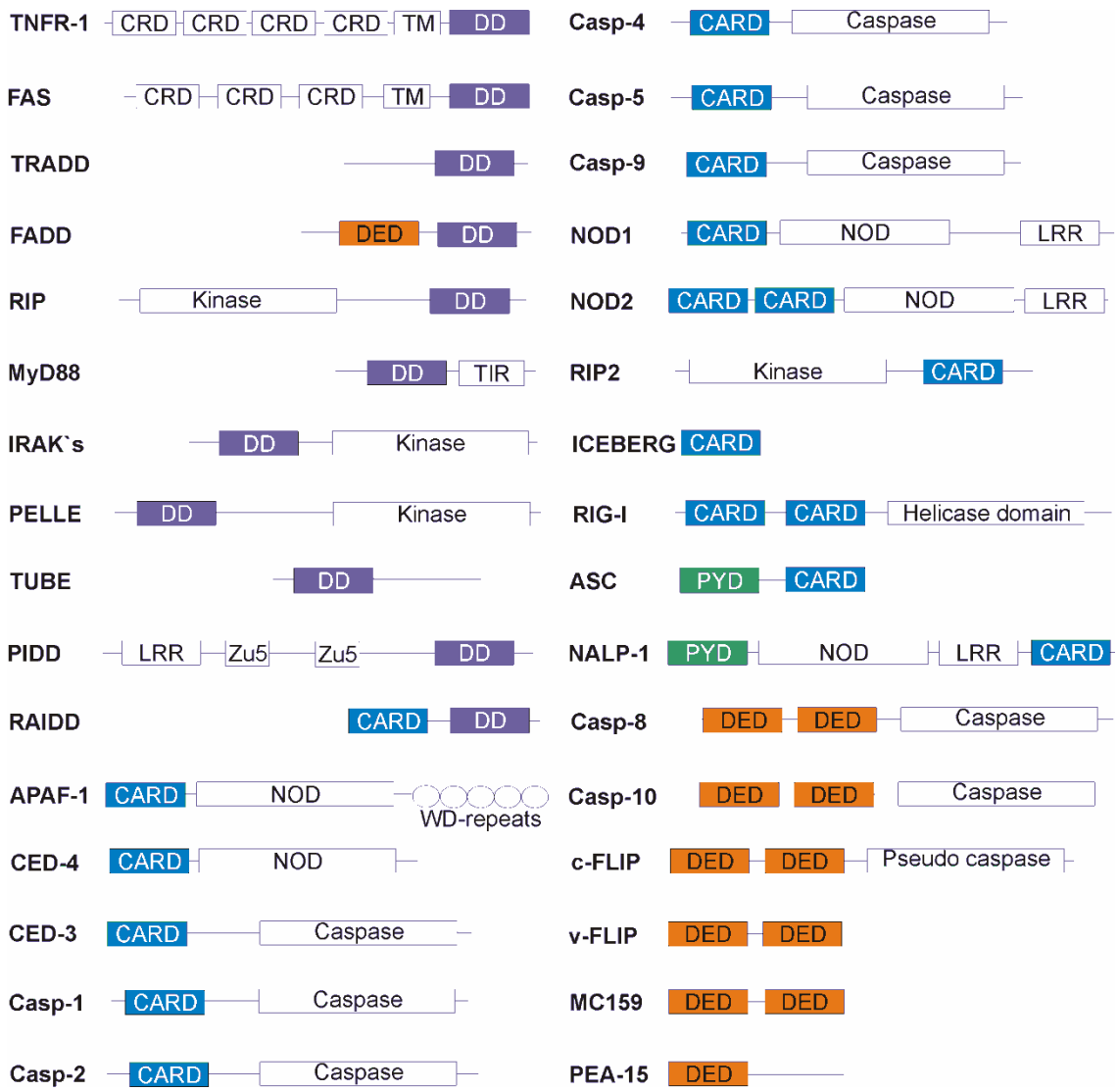


Figure 2_4)
Domain organizations of a selected set of proteins containing the DDF superfamily domains. (Abbreviations in alphabetical order: CARD, caspase activation and recruitment domain; CRD, cysteine-rich domain; DD, death domain; DED, death effector domain; LRR, leucine-rich repeat; NOD, nucleotide-binding oligomerization domain; PYD, pyrin domain; TM, transmembrane; TIR, Toll/interleukin-1 receptor; WD, Trp-Asp repeat, Zu5, ZO-1 and *C.elegans* uncoordinated protein 5. Figure adapted and modified from [8].

Currently, the crystal structures of the homotypic DD interactions of PELLE and TUBE [9] (protein database code 1D2Z), a Drosophila kinase with its adaptor protein, the DD-DD interaction between the PIDD and RAIDD death domains [8] (protein database code 2OF5) a CARD-CARD interaction between Apaf-1 CARD and caspase-9 CARD [10] (protein database code 3YGS) and the DED1:DED2 interaction in the tandem DED of MC

159 a viral FLIP protein from poxvirus [11] (protein database code 2BBR) are known. However, to date, there is no structural data for a homotypic PYD-interaction available.

Two cases of non-homotypic interactions of DDF-superfamily members are known: PEA-15 (phosphoprotein enriched in astrocytes 15) a DED containing protein heterotypically interacts with the structurally unrelated MAP kinase ERK [12]. The second published example is ARC, a CARD containing protein that interacts heterotypically with the death domains of FAS and FADD thereby inhibiting FAS-FADD binding and the assembly formation of the DISC. The same study also shows a non-homotypic interaction between the CARD domain of ARC and the carboxy-terminal regulatory domain of BAX a non-death-fold Bcl-2 family protein [13].

1.1 Intracellular pattern recognition molecules

One important family of proteins containing death domains is the NLR family. This family of intracellular pattern recognition molecules comprises 5 groups, called NLRA, NLRB, NLRC, NLRP and IPAF. The NLRP is the largest group and contains 14 members.

Table 1_4) The human NLR family

NLR	Name	Other names
NLRP	NLRP1	NALP1, DEFCAP, NAC, CARD7, CLR17.1
	NLRP2	NALP2, PYPAF2, NBS1, PAN1, CLR19.9
	NLRP3	NALP3, CRYOPYRIN, PYPAF1, CIAS1, CLR1.1
	NLRP4	NALP4, PYPAF4, PAN2, RNH2, CLR19.5
	NLRP5	NALP5, PYPAF8, MATER, PAN11, CLR19.8
	NLRP6	NALP6, PYPAF5, PAN3, CLR11.4
	NLRP7	NALP7, PYPAF3, NOD12, PAN7, CLR19.4
	NLRP8	NALP8, PAN4, NOD16, CLR19.2
	NLRP9	NALP9, NOD6, PAN12, CLR19.1
	NLRP10	NALP10, PAN5, NOD8, PYNOD, CLR11.1
	NLRP11	NALP11, PYPAF6, NOD17, PAN10, CLR19.6
	NLRP12	NALP12, PYPAF7, Monarch1, RNO2, PAN6, CLR19.3
	NLRP13	NALP13, NOD14, PAN13, CLR19.7
	NLRP14	NALP14, NOD5, PAN8, CLR11.2
NLRC	NOD1	CARD4, CLR7.1
	NOD2	CARD15, IBD1, PSORAS, CLR16.3
	NOD3	CLR16.2
	NOD4	NOD27, CLR16.1
	NOD5	NOD9, CLR11.3
NLRA	CIITA	MHC2TA, C2TA
IPAF	IPAF	CARD12, CLAN, CLR2.1
NLRB	NAIP	BIRC1, CLR5.1

Historically, the family of intracellular pattern recognition molecules was divided into two families, the NALP proteins (NALP1-14) and the NOD proteins (NOD1-5). CIITA, IPAF and NAIP were classified as standalone proteins with similar functions and domains like the NALP and NOD proteins. Recently, this classification scheme has changed. CIITA, IPAF and NAIP proteins now belong to the NLR family, which currently comprises 22 members (Table 1_4) [14].

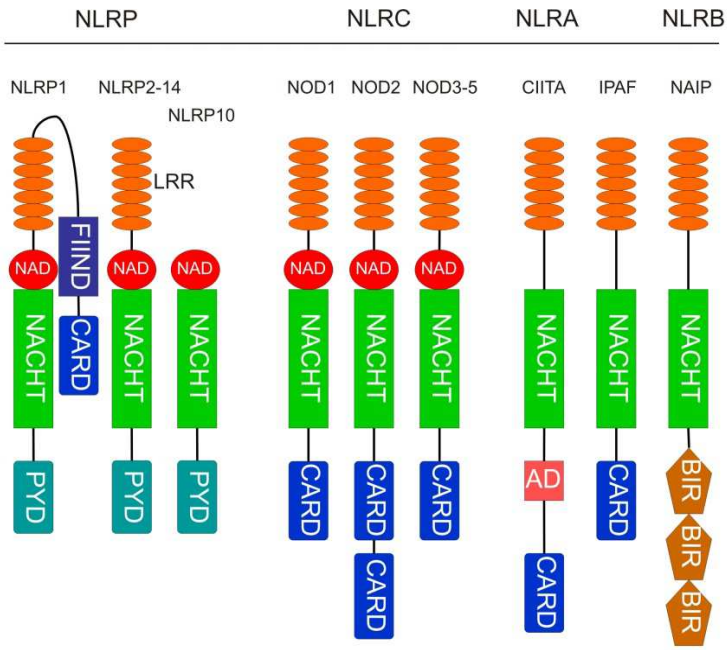


Figure 3_4)
Domain organization of representative NLR. NLR contain three distinct domains: the ligand sensing domain LRR (orange), the oligomerization domain NACHT (green), and the effector domains PYD and CARD (dark green and blue, respectively). CIITA, IPAF and NAIP form the remaining NLR members. Abbreviations: AD: activator domain; BIR: baculovirus inhibitor of apoptosis protein repeat; CIITA: class II transcriptional activator; IPAF: ICE (interleukin-1 converting enzyme) peptidase activating factor; FIIND: function to find; NACHT: NAIP (neuronal apoptosis inhibitor protein), CIITA, HET-E and TP1; NAD: NACHT associated domain; NAIP: neuronal apoptosis inhibitor protein; NOD: nucleotide binding and oligomerization domain. Figure adapted and modified from [15].

Proteins of the NLRP and NLRC group contain at least three distinct domains: An amino-terminal CARD or PYD domain, a nucleotide binding and oligomerization domain (NOD) or NACHT and a carboxy-terminal leucine rich repeat [4]. Furthermore NLRB and NAIP members possess a NACHT associated domain, with so far unknown function. NLRP1 is the only protein of the NLR family with an additional carboxy-terminal CARD domain that is responsible for the recruitment of caspase-5. No functionality has been assigned to the domain named FIIND (Figure 3_4).

The NOD1osome and the NOD2osome are the first reported signalling platforms that detect bacterially derived peptidoglycans [16], [17]. They activate mitogen activated protein kinases (MAPK) via RIP2 and lead to nuclear factor kappa B (NF-κB) release through the recruitment of the IKK complex to the central domain of RIP2 [18], [19], [20] (Figure 4_4). Further studies identify the chemical nature of the compounds and provide evidence that not only the NOD proteins but also NLRP3 are able to sense peptidoglycans [21], [22], [23].

1.2 General overview of inflammasomes

14 different NLRP proteins are known in the human genome (Table 1_4). They contain an amino-terminal PYD domain, a nucleotide binding and oligomerization domain (NACHT) and a carboxy-terminal leucine rich repeat [15]. The only structural information available on NLRP proteins is the nuclear magnetic resonance (NMR) structure of NLRP1-PYD (protein database code 1PN5) [24]. Although structural information is missing, NLRP1, and -3 have been biochemically characterized regarding their state of assembly and their proposed mode of action.

1.2.1 The NLRP1 inflammasome

Acute inflammatory responses are mediated through IL-1β and IL-18 release. Both cytokines are produced as inactive cytoplasmic precursors which are processed into their active forms by caspase-1 [25], [26]. The molecular mechanisms that control the caspase-1 activity rely on a multiprotein complex composed of NLRP1, the protein ASC, caspase-1 and caspase-5 (Figure 4_4) [15], [27], [28].

Mechanistic details have not been published so far, but models describe the signal processing as follows. Upon ligand binding to the LRR of NLRP1, the oligomerization interfaces of the NACHT domain become accessible and NLRP1 oligomerizes. This process is accompanied by large conformational changes and leads to a cluster formation of interaction domains on both termini. The amino-terminal PYD domains of NLRP1 subsequently recruit the adaptor protein ASC through a homotypic PYD-interaction. ASC is a bipartite molecule and comprises a PYD and a CARD domain. ASC, in turn, recruits caspase-1 through a homotypic CARD-interaction. The accessible CARD domain cluster at the carboxy-terminal end of NLRP1 similarly is able to recruit caspase-5 via a homotypic CARD-interaction. The inactive pro-caspase-1 and pro-caspase-5 are transiently brought together, process and activate each other. The exact stoichiometry of this about 700 kD multimeric protein complex is not known. Because this complex regulates pro-inflammatory cytokine activation that cause inflammation symptoms, it was termed "the inflammasome" [29], [30].

The activation of caspase-1 is essential to process inactive IL-1β and IL-18 into their active forms. (Caspase-1 is historically known as Interleukin-1 converting enzyme (ICE)). Once active, IL-1β and IL-18 are secreted and bind to their corresponding IL-1 (IL-1R) and IL-18 receptors (IL-18R) on neighbouring cells [31]. Similarly to Toll-like receptors (TLR), IL-1R and IL-18R possess extracellular LRR and intracellular TIR domains. Since TIR domains are known to recruit adaptor proteins such as MyD88,

which, in turn, activates NF-κB and other signalling cascades, the NLRP activation can be considered a direct link of intracellular pathogen sensing to NF-κB signalling pathways [15].

The ligands that trigger NLRP1 inflammasome activation are unknown. However, extracellular adenosine tri-phosphate, sensed by P2X7 receptors from lipopolysaccaride-primed macrophages, leads to IL-1 cytokine release [32]. In addition, hypotonic stress has been shown to activate NLRP1 by the release of not further specified "host danger signals" [29].

1.2.2 The NLRP3 inflammasome

Similarly to the NLRP1 inflammasome, the NLRP3 inflammasome is a multi-protein complex assembly that controls the processing and activation of IL-1β and IL-18 [33]. In contrast to the NLRP1 inflammasome, the NLRP3 inflammasome comprises ASC, a bipartite molecule termed CARDINAL, which substitutes the two carboxy-terminal FIIND and CARD domains of NLRP1 (absent in NLRP3) and two caspase-1 molecules instead of caspase-1 and caspase-5 (Figure 4_4).

Muramyl-dipeptides, low concentrations of intracellular potassium and potassium channel forming bacterial toxins from *Staphylococcus aureus* or gramicidin from *Bacillus bravis* can trigger NLRP3 inflammasome activation [29], [34], [35]. Uric acid present in tissues as degradation product of purines and calcium pyrophosphate crystals also seem to activate the NLRP3 inflammasome [36], [37]. Moreover, bacterial ribonucleic acid and antiviral compounds are proposed to activate the NLRP3 inflammasome [38].

The stoichiometric composition and the oligomeric state of the ligand sensor are not known for most complexes. The trigger molecule inducing oligomerization for NLRP1 and PIDD is unknown. NLRP3 detects peptidoglycans, NOD detect muramyl-dipeptides and peptidoglycans, TLR sense lipopolysaccharides, viral ribonucleic acid, CpG rich DNA strands and bacterial flagellin each with their leucine rich repeats. The helicase domain of RIG-I detects double stranded ribonucleic acid (Figure 4_4). Upon binding of pathogen associated molecular patterns to their corresponding receptor units, conformational changes expose DD-interaction sites and induce oligomerization analogously to the process as previously described for NLRP1. Adaptor and effector proteins are supposed to be recruited strictly via homotypic DD-interactions.

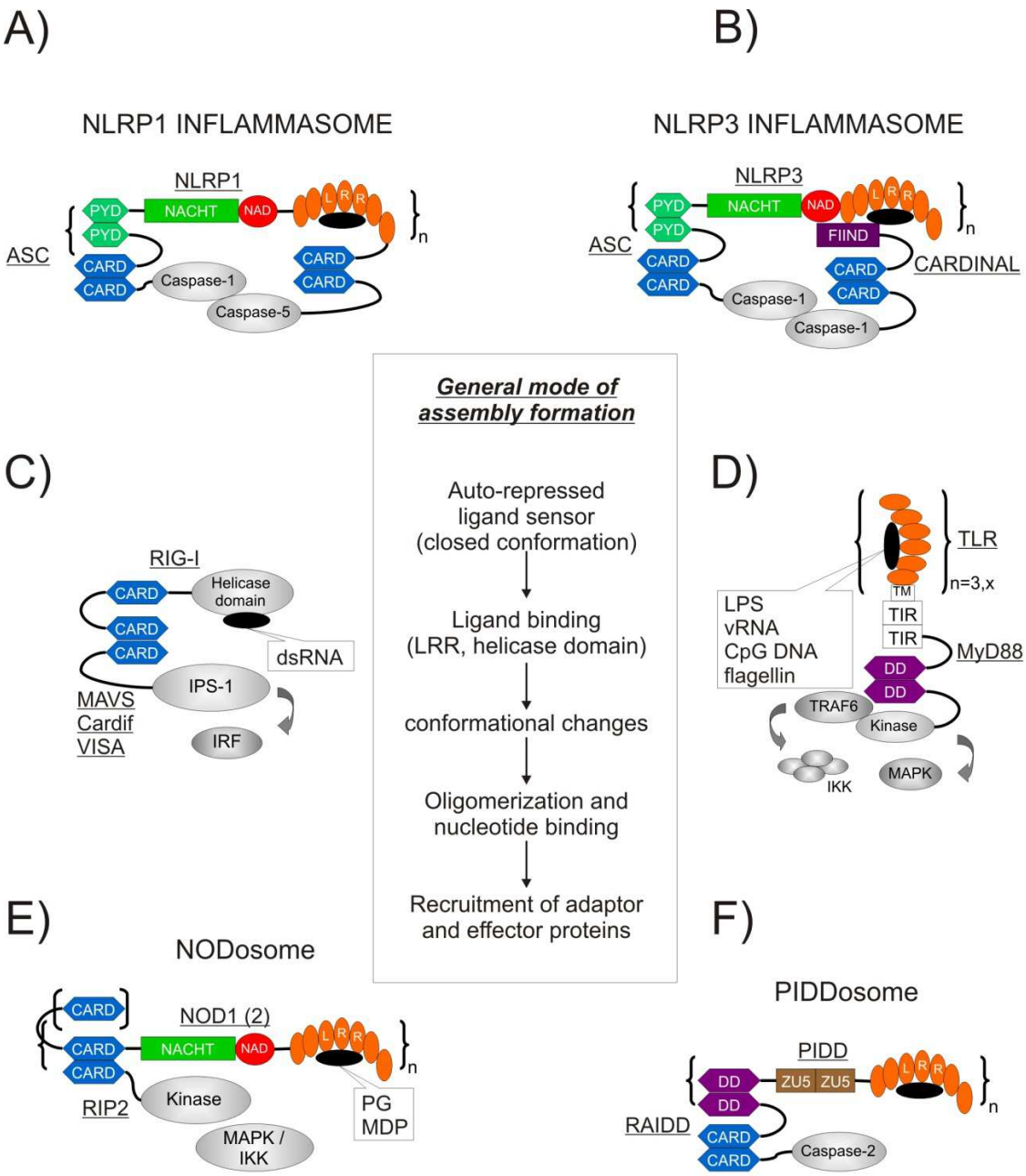


Figure 4_4)
Schematic overview of multiprotein complexes in the innate immune system of A, B) the NLRP1 and NLRP3 inflammasome [30], C) the RIG-I platform [39], D) Toll-like receptors (TLR) [40], E) NODosome [41], and F) the PIDDosome [42]. Abbreviations: dsRNA, double stranded ribonucleic acid; vRNA, viral ribonucleic acid; LPS, lipopolysaccaride; CpG, cytosine-phosphate-guanosine; DNA, desoxyribonucleic acid; PG, peptidoglycan; MDP, muramyl-dipeptide;

1.3 TMS1/ASC

Processes, which maintain the delicate balance between survival and death by death domain containing proteins, can result in diseases when they are imbalanced or even disrupted. In particular the discovery of TMS1 implicated in auto-inflammatory disorders raised interests. Importantly, the loss of expression of TMS1/ASC in breast tumours and other cancers strongly implies its involvement in tumorigenesis.

TMS1/ASC was originally identified by monoclonal antibodies against proteins in an insoluble cytoskeletal fraction of differentiating HL-60 cells [43]. Upon treatment with retinoic acid perinuclear aggregates "specks" were observed giving the protein its name: apoptosis associated speck-like protein containing a CARD domain (ASC). An independent study identified "ASC" in a screen for downstream targets of methylation associated gene silencing. This gene was named TMS1 for Target of Methylation-Induced Silencing-1 (TMS1) and was later found to be identical to ASC [44]. In following sections TMS1/ASC will be named ASC.

1.3.1 ASC in the inflammasome

The bipartite structure of ASC is ideally suited to act as an adaptor molecule by linking some of the NLRP proteins to caspase-1. The amino-terminal PYD domain of NLRP1 and ASC and the carboxy-terminal CARD domain of ASC the CARD domain of caspase-1 can homotypically interact with each other. This adaptor function for ASC has been shown for NLRP1, NLRP3 and NLRP12 [29], [45], [46]. Interestingly, not all NLRP proteins, although very similar in their domain arrangement, require ASC as an adaptor molecule. Dowds et al. show that the PYD domains of NLRP2 and ASC do not interact [47].

1.4 Structural and biochemical investigations of the PYD superfamily

The model of the inflammasomes and interaction modes of the multi-protein assemblies are based on a variety of biochemical studies obtained by immuno-precipitations, yeast two hybrid screens, cellular co-localizations and pull-down experiments. The structures of the NLRP1-PYD and the ASC-PYD domains have been solved by nuclear magnetic resonance (NMR) spectroscopy (protein database codes 1PN5 and 1UCP, respectively) [24], [48]. Interestingly, the helix 3 in both structures is disordered or only partially ordered (Figure 5_4), although the sequences are predicted to assume a six helix bundle. Kaufmann et al. suggested that the helices might play an important role in protein-protein interaction and might represent a significantly different interaction mode compared to other DDF-members [49], [50].

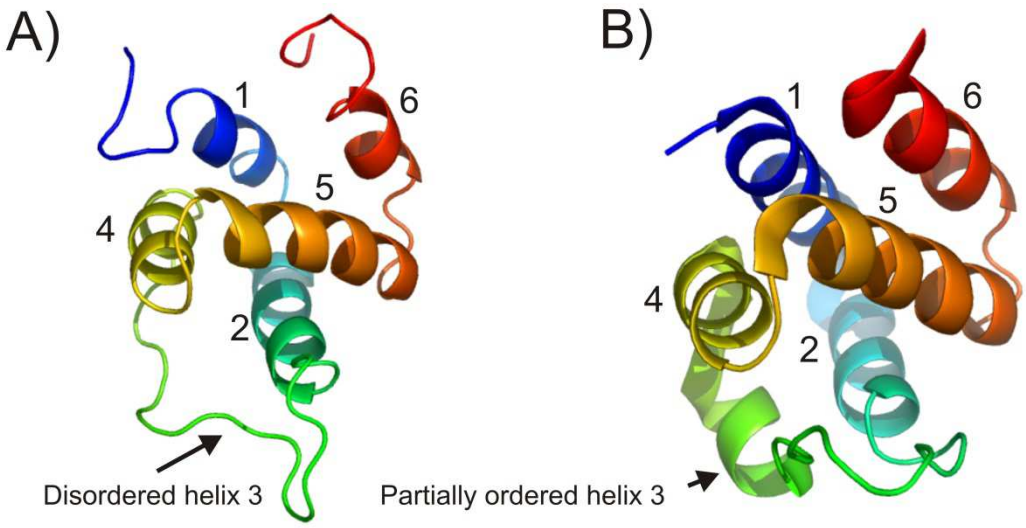


Figure 5_4)
Cartoon representation of PYD domains. A) PYD domain of human NLRP1, protein database code 1PN5; B) PYD domain of ASC, protein database code 1UCP. Helix numbering is indicated; *N*-terminal helix is coloured in blue, *C*-terminal helix is coloured in red.

2. Rationale and aim of the project

The aim of this study is to show a direct homotypic interaction between the PYD domains of NLRP1 and ASC by NMR spectroscopy and to elucidate whether the unstructured loops 3 of both PYD domains might become structured into helices or undergo at least significant structural changes upon dimer formation.

RESULTS

[^{15}N,^{1}H] hetero spin single quantum correlation spectroscopy[10] is a mild procedure that is used elucidate protein-protein interactions. The fundamental requirement for that experiment is a protein that is completely ^{15}N-labelled (^{15}N = heavy nitrogen nuclide). Recording data of the labelled protein in the absence of its interaction partner results in spectrum "A". Recording data of the labelled protein in the presence of its interaction partner results in spectrum "B". Peak shifts might occur in spectrum "B" upon complex formation due to alterations in the local environment of residues that are localized in the interaction interface. The subtraction of "A" from "B" results in a difference spectrum, showing interacting residues (if the peak assignment is known), but the peak assignment is not a prerequisite.

To study the possible complex formation by nuclear magnetic resonance[11] spectroscopy, both domains had to be brought to a condition where they remained monomeric in solution. Extensive solubility tests showed that neither the GB1[12]_NLRP1-PYD nor the ASC-PYD construct stayed in solution at pH values below 6.5 or above pH 4, respectively. Thus, both domains featured incompatible pH conditions to be able to co-exist in solution. Therefore, the ASC-PYD domain was produced in a soluble form as GB1-fusion protein to overcome the insolubility. This construct proved to be soluble and monomeric at low concentrations in the physiological pH range.

3. Protein production and purification

The PYD domain of NLRP1 and the PYD domain of TMS1/ASC were cloned into a modified pET20b+ vector from Novagen, producing a fusion protein with an *N*-terminal GB1-solubility-enhancer tag and a *C*-terminal hexa-histidine sequence (see appendix section).

The PYD domain of ASC was cloned into a pET20 vector (Novagen) which produced the recombinant protein with an *N*-terminal hexa-histidine sequence. Almost uniformly ^{15}N-labelled GB1_NLRP1-PYD was obtained by expressing the protein in *E. coli* BL21 pLysS cells grown in 100 ml minimal medium (M9) supplied with $^{15}NH_4Cl$ as the sole heavy nitrogen source, vitamin mix, solution Q (see Appendix), antibiotics (100 μg/ml ampicillin, 25 μg/ml chloramphenicol), 0.5 % glucose and 1 mM $MgSO_4$. Cells were grown at 37°C to an optical density of OD_{600} = 1.5 – 2 cooled down to 18°C and then induced with 1 mM IPTG (isopropyl-β-D-thiogalacto-pyranoside) for 14 hours. The cells were resuspended in 50 mM Tris, 100 mM NaCl, 3 mM β-mercaptoethanol, 2 mM PMSF (phenylmethylsulphonylfluoride), 1 mM CHAPS (3-[(3-Cholamidopropyl)-dimethylammonio]-1 propanesulfonate, 1 mM imidazole (pH 8.0 at RT), lyzed with a French press. The protein was purified by IMAC (immobilized metal affinity chromatography) on Ni^{2+}-loaded nitrilo-triacetic acid (NTA) matrix from Qiagen and by

[10] [15N,1H] hetero spin single quantum correlation spectroscopy; HSQC
[11] Nuclear magnetic resonance; NMR
[12] B1-domain of protein G; GB1

size exclusion chromatography on a S75 26/60 column (Amersham, Pharmacia Biotech) thereby getting rebuffered into 20 mM Tris (2-amino-2-hydroxymethyl-propane-1,3-diol, 100 mM NaCl (sodium chloride), 0.2 mM CHAPS, 0.1 mM EDTA (ethylenediamine tetraacetic acid), 10 mM DTT (dithiothreitol) pH 7.5 at 4°C. This buffer was used for circular dichroism spectroscopy[13] and dynamic light scattering experiments.

Unlabelled GB1_ASC-PYD was obtained by expressing the protein in *E. coli* BL21 pLysS cells grown in 500 ml Luria Bertani cultures using the same conditions as described above.

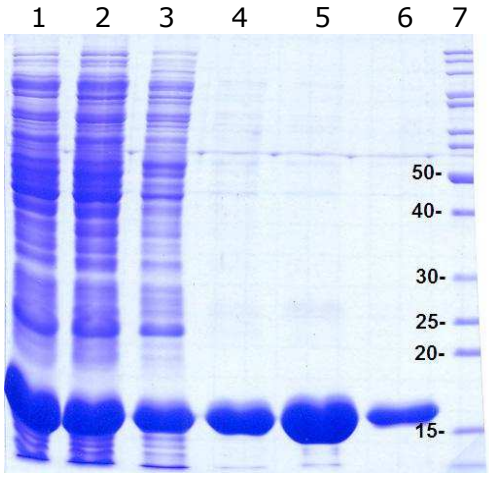


Figure 6_4)
15% SDS-PAGE of IMAC purified GB1-ASC-PYD; lane 1: supernatant; lane 2: flow through; lane 3: wash; lanes 4-6: elution fractions; lane 7: marker, molecular weights indicated.

The protein was purified by IMAC and by size exclusion chromatography on a S75 26/60 column (Amersham, Pharmacia Biotech) thereby getting rebuffered in 20 mM Tris, 500 mM NaCl, 0.2 mM CHAPS, 0.1 mM EDTA, 10 mM DTT pH 7.5 at 4°C (Figure 6_4). This buffer was used for CD-spectroscopy and dynamic light scattering experiments. Prior to NMR measurements the proteins were dialyzed for 24 hours against 50 mM Na_2HPO_4/NaH_2PO_4, pH 7.0, 5% D_2O.

3.1 Mass analysis and protein purity

The calculated relative molecular masses of GB1_ASC-PYD and uniformly labelled ^{15}N GB1_NLRP1-PYD are 17493.8 Da and 20784.8 Da, respectively (www.expasy.org; protparam). The quadrupole time of flight mass spectra electron spray ionization[14] confirmed the calculated molecular masses (Figure 8_4). The labelling degree of GB1_NLRP1-PYD was estimated to be greater than 95%, sufficient for adequate NMR

[13] Circular dichroism spectroscopy; CD spectroscopy
[14] Quadrupole time of light mass spectra electron spray ionization; QTOF MS ESI

measurements (Hiller S., personal comment). Since ESI spectra display the relevant mass peaks only within a mass window, GB1_ASC-PYD was also analyzed on a C8 column by high pressure liquid chromatography[15] to confirm its purity (Figure 7_4).

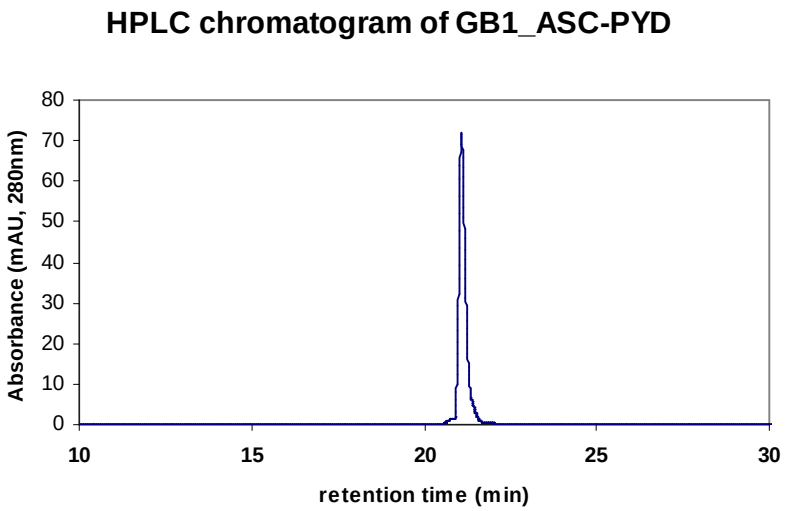


Figure 7_4)
HPLC chromatogram of purified GB1_ASC-PYD. Purified on a C8 column using a gradient of 0-40% acetonitrile and 0.1% TFA (trifluoro-acetic acid)

Previous studies have shown that the NLRP1-PYD construct is not soluble without the GB1-solubility-enhancer tag in a variety of buffer conditions (including 10% glycerol and sucrose, high and low salt concentrations) from pH 3 to pH 9 (Kohl, A., personal comment). The GB1_NLRP1-PYD construct had improved solubility behaviour and remained in solution above pH 6.5 without stabilizing additives such as glycerol and sucrose or divalent salt ions.

The NMR structure of ASC-PYD with an *N*-terminal hexa-histidine sequence was recorded at pH 3.7, where it is stable and monomeric in solution [48]. However, the domain aggregated and precipitated in buffered systems above pH 4. Because ASC-PYD is not soluble above pH 4 and GB1_NLRP-PYD is not soluble below pH 6.5 it was impossible to analyze the protein-protein interaction by NMR spectroscopy or any other method. Therefore ASC-PYD was produced with a GB1-solubility-enhancer tag. This successfully prevented aggregation at a neutral pH. However, GB1_ASC-PYD had the tendency to precipitate over time (days), at concentrations above 1 – 2 mg/ml at neutral pH, and readily precipitated within minutes when precipitated protein was already present. Both GB1-tagged domains proved to be monomeric in solution at concentrations around 1 mg/ml. This state was stable and the concentrations were sufficient for overnight [^{15}N,^{1}H] HSQC measurements.

[15] High pressure liquid chromatography; HPLC

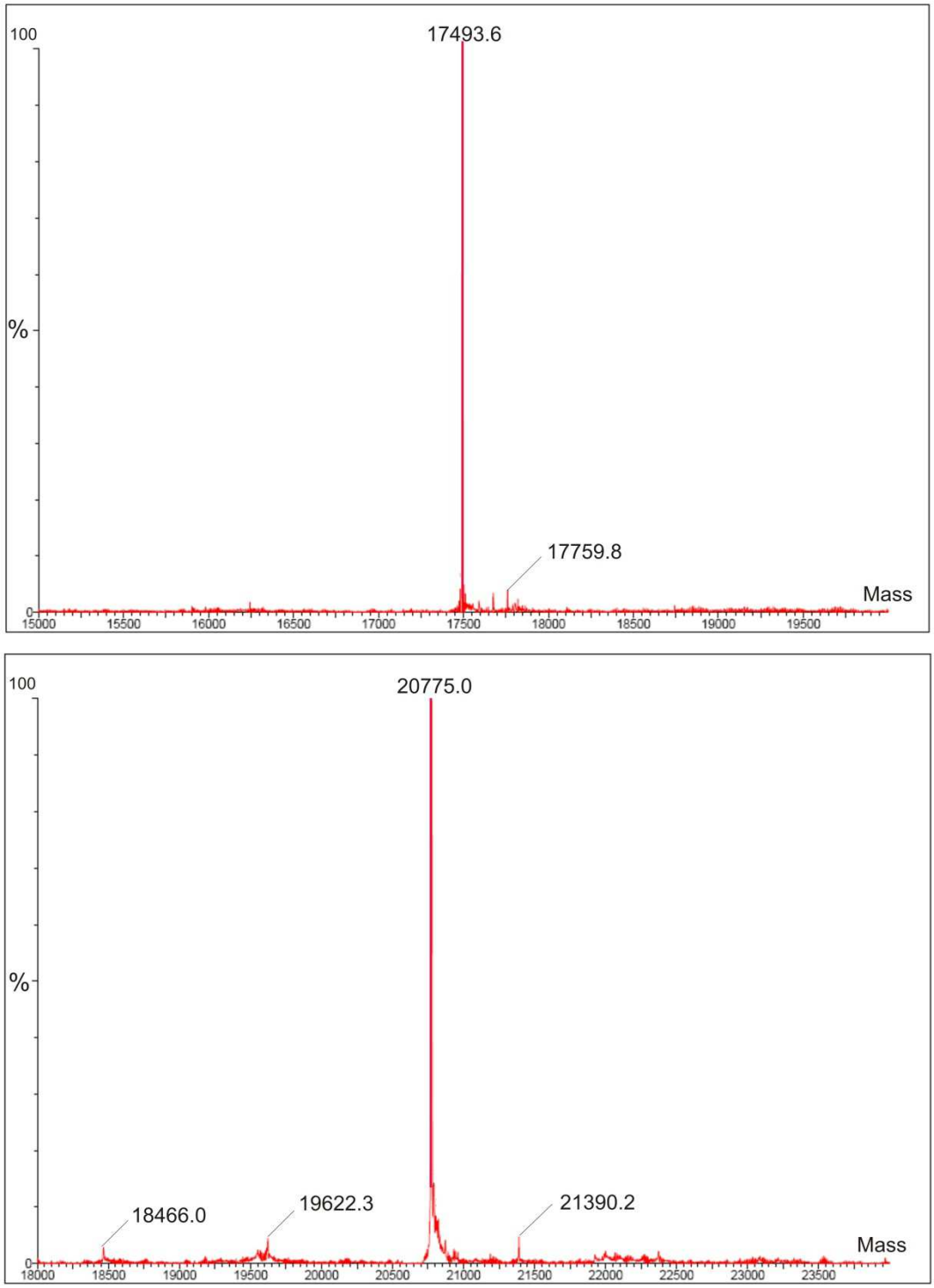


Figure 8_4)
Quadrupole time of flight mass spectrum and electrospray ionization of the unlabelled GB1_ASC-PYD domain (upper panel) and ^{15}N labelled GB1_NLRP1-PYD domain (lower panel). The experimentally determined masses of 17493.6 Da and 20775.0 Da are in agreement with the calculated masses of 17493.8 Da and 20784.8 Da. The x-axis indicates the mass, the y-axis the percentage of intensity, respectively. The measurements were performed at the Functional Genomics Centre Zurich (FGCZ).

3.2 CD spectroscopy

The GB1_NLRP1-PYD protein is correctly folded in a buffer containing 20 mM Tris, 100 mM NaCl, 10 mM DTT, 1 mM CHAPS, 0.1 mM EDTA pH 7.5 (at room temperature) (Kohl, A., personal comment). The CD spectrum of ASC-PYD was recorded at pH 4 in 20 mM NaH_2PO_4 showing two absorption minima at 223 nm and 208 nm reminiscent of a protein with a high α-helical content. The CD spectrum of GB1_ASC-PYD was recorded at pH 7.5 in 20 mM Tris 100 mM NaCl and similarly showed a α-helical curve, with absorption minima at 223 nm and 208 nm (Figure 9_4).

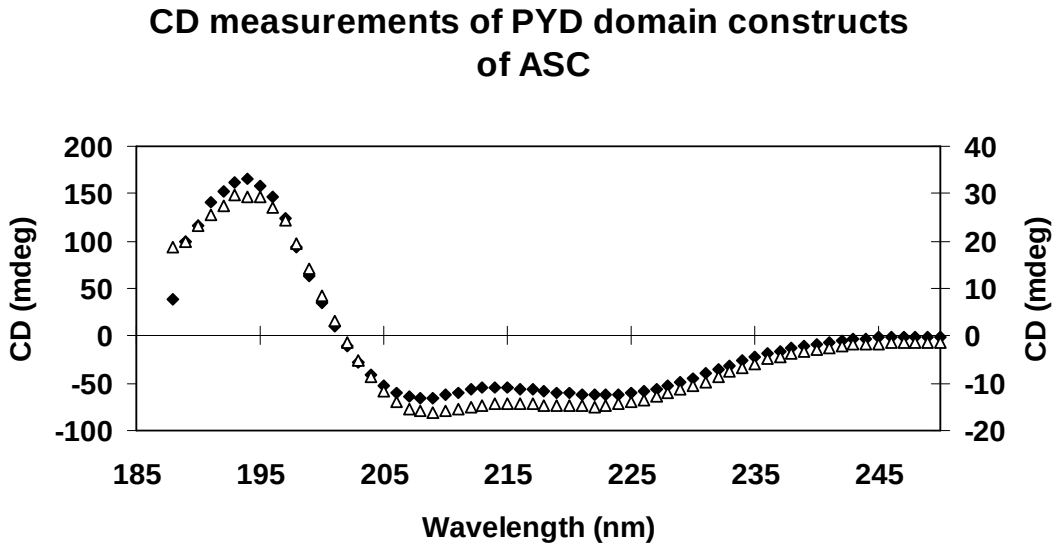


Figure 9_4)
Overlaid circular dichroism spectra of the ASC-PYD domain (◆) recorded at pH 4.0 left scale (y-axis) and the GB1_ASC-PYD domain (△) recorded at pH 7.5 right scale (y-axis)

3.3 NMR measurement

To structurally determine the proposed homotypic PYD-interaction between GB1_ASC-PYD and GB1_NLRP1-PYD, the [^{15}N,^{1}H]-HSQC run was kindly performed by Sebastian Hiller (group of Prof. Dr. K. Wüthrich, ETH, Zürich).

A solution containing 200 μM GB1_NLRP1-PYD was mixed with a solution containing 80 μM GB1_ASC-PYD. The two samples had been dialyzed for 24 hours in the same buffer (50 mM NaH_2PO_4/Na_2HPO_4, buffered at pH 7.0). The GB1_NLRP1-PYD fusion protein was uniformly ^{15}N-labelled; the GB1_ASC-PYD fusion protein was unlabelled. One dimensional ^{1}H-NMR spectra verified that both proteins were still in solution and had not aggregated in the mixed sample. The [^{15}N,^{1}H]-HSQC spectra of GB1_NLRP1-PYD and together with its putative binding partner GB1_ASC-PYD were identical (Figure 10_4).

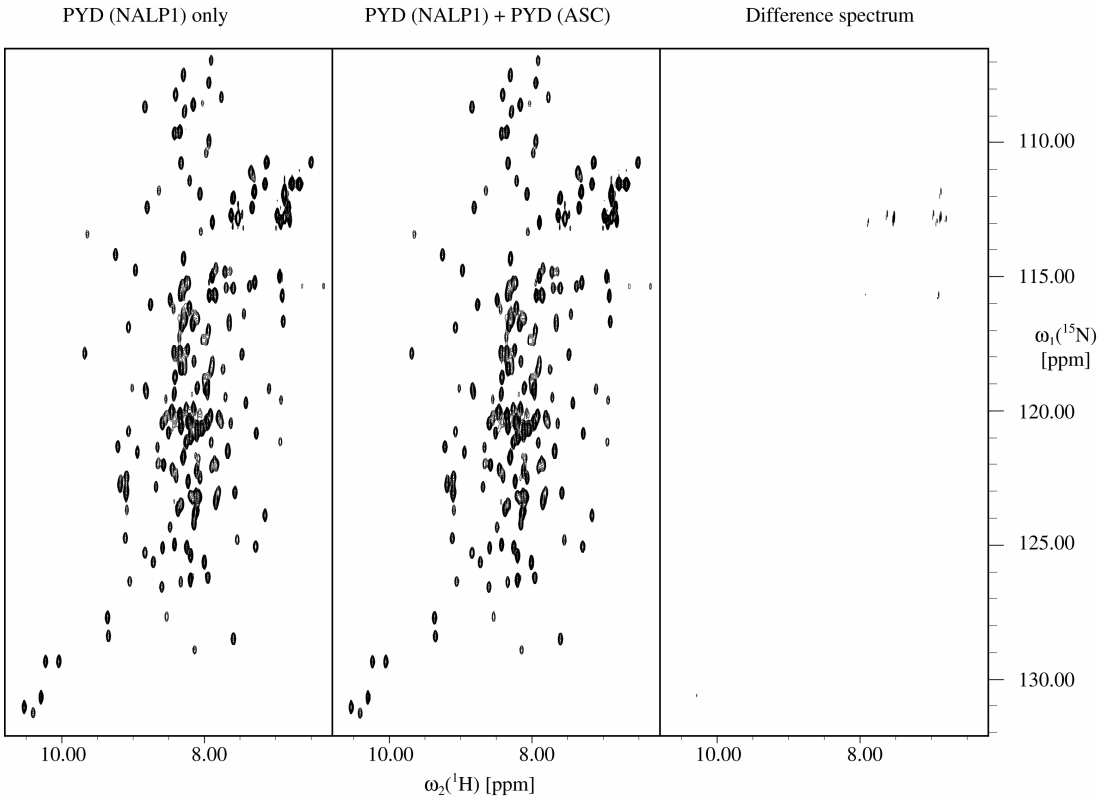


Figure 10_4)
[^{15}N,^{1}H] HSQC spectra; left: spectrum with uniformly labelled ^{15}N-GB1_NLRP1-PYD; middle: spectrum of a mixture of uniformly labelled ^{15}N-GB1_NLRP1-PYD and unlabelled GB1_ASC-PYD; right: difference spectrum of ^{15}N-GB1_NLRP1-PYD minus ^{15}N-GB1_NLRP1-PYD mixed with unlabelled GB1_ASC-PYD. The difference spectrum does not show significant peak shifts, which would occur upon complex formation.

Based on the known protein concentrations at the beginning of the experiment, and the calculated signal intensity of the NMR spectra, the calculated concentration of a complex between the two PYD domains was below 5% (based on the concentration of uncomplexed GB1_NLRP1-PYD). This allows the calculation of a dissociation constant (K_D), which, taking the known parameters into account is 1.3 mM (see equation I). The concentration of 1.3 mM is above physiological relevance. From this result, it can be concluded, that the two PYD domains, both in construct with a GB1 fusion protein, do not interact under the buffer conditions used for this experimental setup.

$$K_D < \frac{[NLRP1]\cdot[ASC]}{[NLRP1-ASC]} = \frac{190\cdot 70}{10}\mu M = 1.3mM$$

equation I

DISCUSSION

Despite the importance of death domains in inflammatory signalling and regulatory mechanisms of the innate immune system, only few structures (X-ray and NMR) have been solved in recent years. The overall low number of structures of DDF proteins can reasonably be explained with solubility difficulties. As shown in previous sections, the ASC-PYD and the NLRP1-PYD domain could only be kept in solution at either non-physiological pH values or with solubility enhancing fusion proteins. As a consequence of lacking structural data, information about death domain interactions is limited to few representatives of each DDF domain.

Generally, death domains are small in size and well-expressed in bacterial expression systems, but are prone to aggregation, and mainly found in either insoluble inclusion bodies, globular aggregates (specks) or filamentous structures [43], [50]. As described before, the NLRP1-PYD domain could not be stabilized and kept in solution in a variety of buffers over a broad pH range. The ASC-PYD domain was only soluble in a non-physiological acidic buffer.

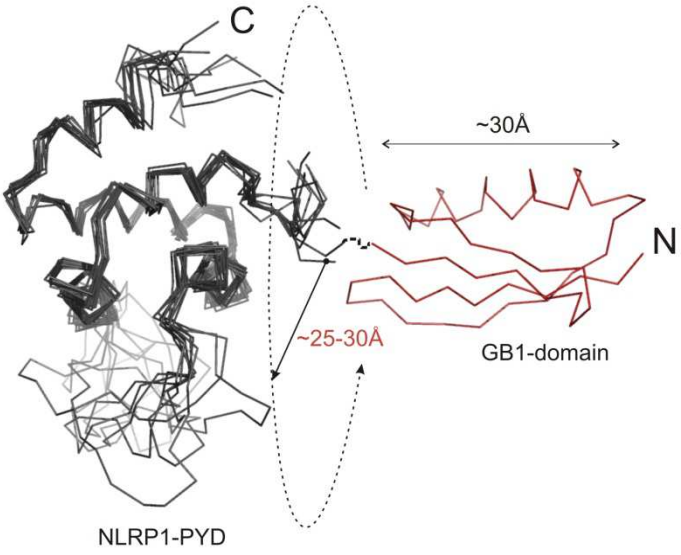


Figure 11_4)
GB1_NLRP1-PYD fusion protein; The GB1-domain depicted in red (protein database code 2J52) was positioned close to the "amino-terminus" of the NLRP1-PYD structure (10 average structures shown in black) (protein database code 1PN5). Two linker residues between the domains are depicted as dotted line. The estimated length of the GB1-domain is drawn as a circle indicating the blocked area on the NLRP1-PYD domain.

The crucial point of the experimental setup was the amino-terminal modification with a solubility enhancing GB1-tag. This modification allowed the two fusion proteins to co-exist in a monomeric state in a physiological buffer. Construct truncations or elongations such as solubility tags (GB1 domain or even fusion proteins), detection tags or

purification tags, always bear the risk to completely abrogate possible interactions. Figure 11_4 shows for this experimental setup, where the *N*-terminal GB1-fusion protein would be located with respect to the disordered helix 3 of the NLRP1-PYD domain.

It has been shown that short detection tags such as vesicular stomatitis virus (VSV-tag), Flag-tag, myelocytomatosis virus (Myc-tag) or hexa-histidine-tag do not affect domain-domain interactions. For example: VSV-tagged full length ASC, co-expressed with FLAG-tag PYD of ASC (without CARD) was immuno-precipitated. Moreover, the same study showed that full length VSV-tagged ASC interacts with the FLAG-tagged NLRP1-PYD domain [51]. A summary of published results about death domain interactions is depicted in Figure 12_4.

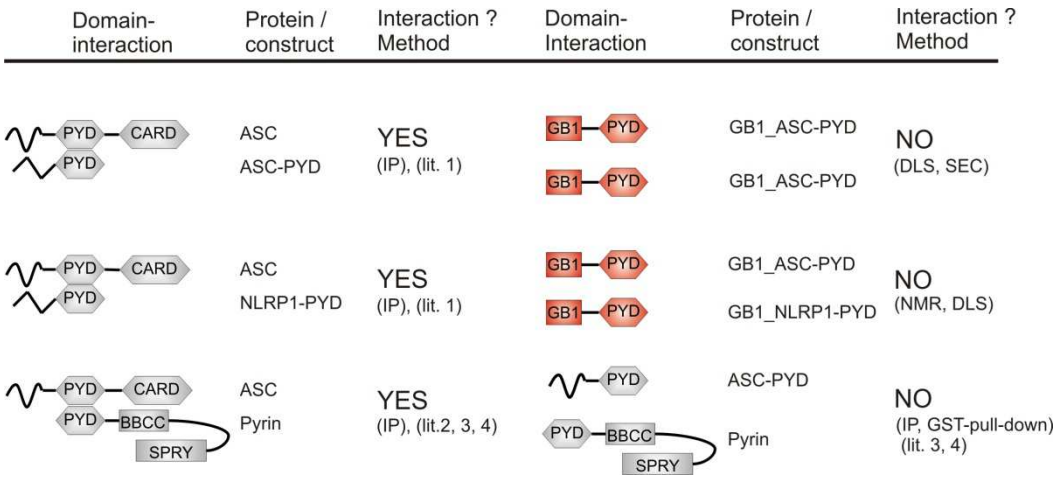


Figure 12_4)
Schematic drawing of investigated protein-protein interactions. Abbreviations used: IP: immuno-precipitation, DLS: dynamic light scattering, SEC: size exclusion chromatography, NMR: nuclear magnetic resonance. Constructs (with N-terminal waved tags): VSV-ASC / FLAG-PYD (lit.1); Myc-ASC or Myc-ASC-PYD / FLAG-NLRP1-PYD / VSV-pyrin (lit. 1 and 2); Myc-ASC-PYD (lit. 3 and 4). Indicated literature references: 1) Martinon et al., 2001; 2) Papin et al., 2007; 3) Srinivasula et al., 2002 4) Richards et al., 2001.

However, the [^{15}N,^{1}H] HSQC experiment showed that the GB1-tagged NLRP1-PYD domain and the GB1-tagged ASC-PYD domain did not interact under given conditions. Several reasons might account for this experimental finding.

It cannot be excluded that the GB1 domain in contrast to the VSV-tag, Flag-tag or Myc-tag does not negatively influence domain-domain interactions. Whereas the detection tags are highly flexible and short amino acid sequences, the GB1 domain is a stable entity comprising 56 amino acids. Thus, steric hindrance might be a major criterion preventing interactions. Furthermore, due to important regulatory mechanisms and consequences of the release of IL-1β, the protein interactions in the native

inflammasome must be tightly controllable by the cell. The inflammasome complex must remain transient, built upon weak interactions, during its processing state, maybe too weak to be measured as single domain constructs. Intracellular conditions are of utmost importance regulating the complex assembly and disassembly *in vivo*. The NMR experiment was performed in low salt conditions (50 mM NaH_2PO_4/Na_2HPO_4, at pH 7.0) without further additives, a condition that is far from a cytosolic milieu.

Nevertheless, this study showed the necessary measures to bring the two PYD domains together in solution. Martinon et al., 2001, shows in a different experiment that the PYD domains of NLRP1 and ASC interact [51]. Thus, only minor changes in the experimental setup might be required to be able to show the NLRP1-PYD – ASC-PYD interaction by NMR and to elucidate their interaction mode.

REFERENCES

1.) Barton, G.M. and R. Medzhitov, Toll-like receptors and their ligands. Curr Top Microbiol Immunol, 2002. 270: p. 81-92.

2.) Inohara, N. and G. Nunez, The NOD: a signaling module that regulates apoptosis and host defense against pathogens. Oncogene, 2001. 20(44): p. 6473-81.

3.) Philpott, D.J., S.E. Girardin, and P.J. Sansonetti, Innate immune responses of epithelial cells following infection with bacterial pathogens. Curr Opin Immunol, 2001. 13(4): p. 410-6.

4.) Tschopp, J., F. Martinon, and K. Burns, NALPs: a novel protein family involved in inflammation. Nat Rev Mol Cell Biol, 2003. 4(2): p. 95-104.

5.) Chamaillard, M., et al., Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. Cell Microbiol, 2003. 5(9): p. 581-92.

6.) Creagh, E.M., H. Conroy, and S.J. Martin, Caspase-activation pathways in apoptosis and immunity. Immunol Rev, 2003. 193: p. 10-21.

7.) Medzhitov, R. and C. Janeway, Jr., Innate immune recognition: mechanisms and pathways. Immunol Rev, 2000. 173: p. 89-97.

8.) Park, H.H., et al., Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex. Cell, 2007. 128(3): p. 533-46.

9.) Xiao, T., et al., Three-dimensional structure of a complex between the death domains of Pelle and Tube. Cell, 1999. 99(5): p. 545-55.

10.) Qin, H., et al., Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature, 1999. 399(6736): p. 549-57.

11.) Yang, J.K., et al., Crystal structure of MC159 reveals molecular mechanism of DISC assembly and FLIP inhibition. Mol Cell, 2005. 20(6): p. 939-49.

12.) Hill, J.M., et al., Recognition of ERK MAP kinase by PEA-15 reveals a common docking site within the death domain and death effector domain. Embo J, 2002. 21(23): p. 6494-504.

13.) Nam, Y.J., et al., Inhibition of both the extrinsic and intrinsic death pathways through nonhomotypic death-fold interactions. Mol Cell, 2004. 15(6): p. 901-12.

14.) Ting, J.P.-Y., Lovering, R.C., Alnemri, E.S., Bertin, J., Boss, J.M., Davis, B.K., Flavell, R.A., Girardin, S.E., Godzik, A., Harton, J.A., Hoffman, H.M., Hugot, J-P., and N. Inohara, MacKenzie, A., Maltais, L.J., Nunez, G., Ogura, Y., Otten, L.A., Philpott, D., Read, J.C., Reith, W., Schreiber, S., Steimle, V., Ward, P.A., The NLR Gene Family: A Standard Nomenclature. Immunity, 2008. 28: p. 285-287.

15.) Martinon, F. and J. Tschopp, NLRs join TLRs as innate sensors of pathogens. Trends Immunol, 2005. 26(8): p. 447-54.

16.) Inohara, N., et al., Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. J Biol Chem, 1999. 274(21): p. 14560-7.

17.) Ogura, Y., et al., Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. J Biol Chem, 2001. 276(7): p. 4812-8.

18.) Inohara, N., et al., An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. J Biol Chem, 2000. 275(36): p. 27823-31.

19.) Inohara, N., et al., Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem, 2003. 278(8): p. 5509-12.

20.) Viala, J., et al., Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. Nat Immunol, 2004. 5(11): p. 1166-74.

21.) Ting, J.P. and B.K. Davis, CATERPILLER: a novel gene family important in immunity, cell death, and diseases. Annu Rev Immunol, 2005. 23: p. 387-414.

22.) Martinon, F., et al., Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. Curr Biol, 2004. 14(21): p. 1929-34.

23.) Girardin, S.E., et al., Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. J Biol Chem, 2003. 278(11): p. 8869-72.

24.) Hiller, S., et al., NMR structure of the apoptosis- and inflammation-related NALP1 pyrin domain. Structure, 2003. 11(10): p. 1199-205.

25.) Thornberry, N.A., et al., A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. Nature, 1992. 356(6372): p. 768-74.

26.) Dinarello, C.A., Interleukin-1 beta, interleukin-18, and the interleukin-1 beta converting enzyme. Ann N Y Acad Sci, 1998. 856: p. 1-11.

27.) Srinivasula, S.M., et al., The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. J Biol Chem, 2002. 277(24): p. 21119-22.

28.) Agostini, L., et al., NALP3 forms an IL-1beta-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. Immunity, 2004. 20(3): p. 319-25.

29.) Martinon, F., K. Burns, and J. Tschopp, The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Mol Cell, 2002. 10(2): p. 417-26.

30.) Martinon, F. and J. Tschopp, Inflammatory caspases: linking an intracellular innate immune system to autoinflammatory diseases. Cell, 2004. 117(5): p. 561-74.

31.) Suzuki, N., S. Suzuki, and W.C. Yeh, IRAK-4 as the central TIR signaling mediator in innate immunity. Trends Immunol, 2002. 23(10): p. 503-6.

32.) Pelegrin, P., C. Barroso-Gutierrez, and A. Surprenant, P2X7 receptor differentially couples to distinct release pathways for IL-1beta in mouse macrophage. J Immunol, 2008. 180(11): p. 7147-57.

33.) Meylan, E., J. Tschopp, and M. Karin, Intracellular pattern recognition receptors in the host response. Nature, 2006. 442(7098): p. 39-44.

34.) Perregaux, D. and C.A. Gabel, Interleukin-1 beta maturation and release in response to ATP and nigericin. Evidence that potassium depletion mediated by these agents is a necessary and common feature of their activity. J Biol Chem, 1994. 269(21): p. 15195-203.

35.) Walev, I., et al., Potassium-inhibited processing of IL-1 beta in human monocytes. Embo J, 1995. 14(8): p. 1607-14.

36.) Shi, Y., J.E. Evans, and K.L. Rock, Molecular identification of a danger signal that alerts the immune system to dying cells. Nature, 2003. 425(6957): p. 516-21.

37.) Martinon, F., et al., Gout-associated uric acid crystals activate the NALP3 inflammasome. Nature, 2006. 440(7081): p. 237-41.

38.) Kanneganti, T.D., et al., Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. Nature, 2006. 440(7081): p. 233-6.

39.) Meylan, E. and J. Tschopp, Toll-like receptors and RNA helicases: two parallel ways to trigger antiviral responses. Mol Cell, 2006. 22(5): p. 561-9.

40.) Chen, K., et al., Toll-like receptors in inflammation, infection and cancer. Int Immunopharmacol, 2007. 7(10): p. 1271-85.

41.) Rosenstiel, P., A. Till, and S. Schreiber, NOD-like receptors and human diseases. Microbes Infect, 2007. 9(5): p. 648-57.

42.) Tinel, A. and J. Tschopp, The PIDDosome, a protein complex implicated in activation of caspase-2 in response to genotoxic stress. Science, 2004. 304(5672): p. 843-6.

43.) Masumoto, J., et al., ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells. J Biol Chem, 1999. 274(48): p. 33835-8.

44.) Conway, K.E., et al., TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers. Cancer Res, 2000. 60(22): p. 6236-42.

45.) Manji, G.A., et al., PYPAF1, a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF-kappa B. J Biol Chem, 2002. 277(13): p. 11570-5.

46.) Wang, L., et al., PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates activation of NF-kappa B and caspase-1-dependent cytokine processing. J Biol Chem, 2002. 277(33): p. 29874-80.

47.) Dowds, T.A., et al., Regulation of cryopyrin/Pypaf1 signaling by pyrin, the familial Mediterranean fever gene product. Biochem Biophys Res Commun, 2003. 302(3): p. 575-80.

48.) Liepinsh, E., et al., The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition. J Mol Biol, 2003. 332(5): p. 1155-63.

49.) Kaufmann, M., et al., Identification of a basic surface area of the FADD death effector domain critical for apoptotic signaling. FEBS Lett, 2002. 527(1-3): p. 250-4.

50.) Richards, N., et al., Interaction between pyrin and the apoptotic speck protein (ASC) modulates ASC-induced apoptosis. J Biol Chem, 2001. 276(42): p. 39320-9.

51.) Martinon, F., K. Hofmann, and J. Tschopp, The pyrin domain: a possible member of the death domain-fold family implicated in apoptosis and inflammation. Curr Biol, 2001. 11(4): p. R118-20.

APPENDIX

RIBOSOME DISPLAY AGAINST SELECTED CASPASES AND PROTOCOLS

RIBOSOME DISPLAY AGAINST SELECTED CASPASES

Caspase-1, -2, -3, -5 and -8 were used as target proteins for the initial ribosome display[1] selection with a N2C messenger RNA library (provided by Molecular Partners). The selection of caspase-1 binders was not successful, because the protein was unstable in any chosen buffer. A pool of caspase-2 binders (N3C) is currently investigated. Several binders against caspase-3 have been selected and characterized regarding the formation of protein-protein complexes and their inhibitory activity. However, further biophysical characterization is still ongoing. Due to the limited stability of caspase-5 further RD selection rounds are necessary. Nevertheless, a first selection trial yielded a few positive clones, including several caspase-8 binders, which also need further characterization. Preliminary binding experiments were not conclusive so far. Reasonable binding results were only obtained with caspase-3 binders, and data will be presented in the following sections.

Enzyme linked immuno sorbent assay[2]

The detailed ELISA and RD protocols are given in the protocol section of the appendix.

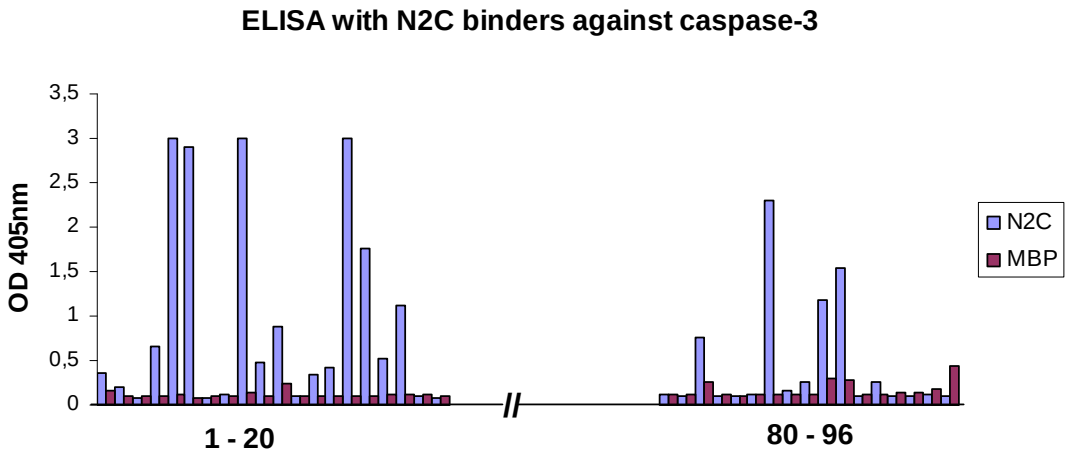


Figure 1_A)
ELISA readout from two 96well plates coated with caspase-3 and MBP: Blue bars represent binding of N2C ankyrin binders against caspase-3 adjacent dark pink bars represent the background signal of that particular binder against MBP. This data has been acquired three days after the start of the colour development, indicating that selected binders do not give background signals for MBP.

After four RD rounds, the DNA fragments encoding N2C library members for putative caspase-3 binders were cloned into expression vectors (pQE-30), transformed into *E. coli* XL-1 blue cells and streaked out onto LB agar plates supplied with appropriate antibiotics

[1] Ribosome display; RD
[2] Enzyme linked immuno sorbent assay; ELISA

(ampicillin and tetracycline). Single colonies were picked and separately expressed in 96-well plates. After cell lysis, the crude extracts containing N2C binders were transferred onto caspase-3 / maltose binding protein (MBP) coated plates. MBP was used as target protein for pre-panning selections during RD rounds. For every well coated with caspase-3, a second well was coated with MBP serving as negative control for caspase-3 binders, and as positive control for a previously selected MBP binder. All binders were expressed with an amino-terminal hexa-histidine tag, which was recognized by the primary antibody. The signal readout originated from an alkaline phosphatase coupled secondary antibody that turned 4-nitrophenyl phosphate into a yellow coloured product (4-nitrophenolate; detected at 405 nm wavelength). The signal followed over time (30 min intervals, and 1-2 days), indicated the quality of the background binding (Figure 1_A).

Size exclusion chromatography (SEC)

Caspase-3 and the selected binders against caspase-3, chosen from the previous ELISA experiment, were separately analyzed by size exclusion chromatography on a S200 (Sephadex, 10/300) column (Pharmacia) to obtain the retention times each in the unbound form. The binders and caspase-3 were subsequently allowed to form a complex by mixing each individual binder with caspase-3 in a 2.2:1 molar excess in 150 µl volume. This mixture incubated on ice for 2 hours, was centrifuged for 5 min and then injected. Calculated theoretical molecular masses of selected N2C binders against caspase-3 are given in Table 1_A.

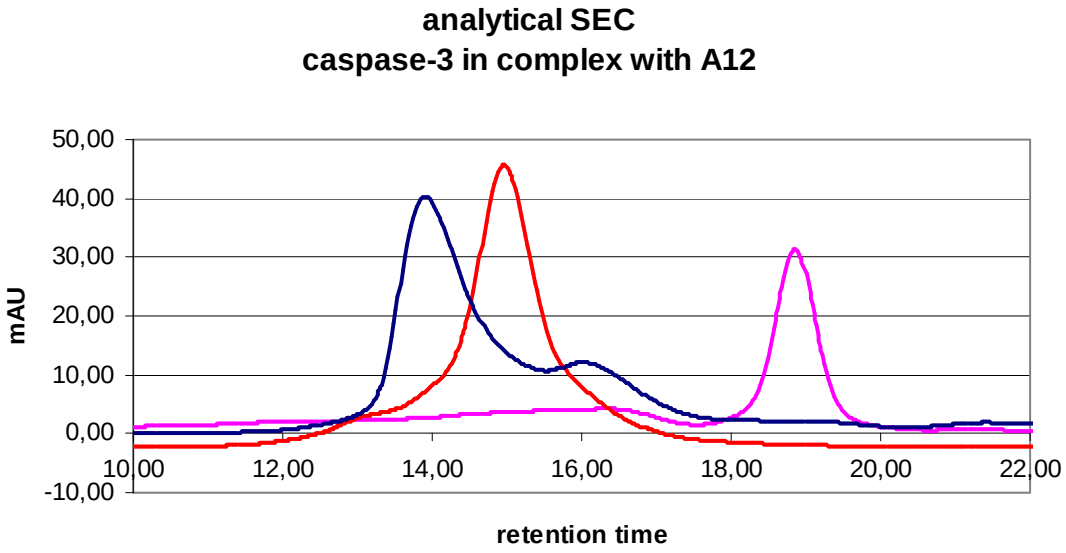


Figure 2_A)
Overlaid chromatograms of the N2C binder (A12) against caspase-3 (pink profile), caspase-3 (red profile) and caspase-3 in complex with A12 (blue profile). The peaks (retention time) correspond to molecular masses of 15 kD for a single N2C ankyrin, 60 kD for the caspase-3 and 93 kD for the complex. The loading volume was 100 µl; flow rate: 0.5 ml/min; buffer: 20 mM Tris pH 8.0, 150 mM NaCl; temperature 4°C, S200 (10/300) Sephadex column (Pharmacia).

Caspase-3 is a stable dimer in solution and displays a molecular mass of 60 kD according to its retention time in size exclusion chromatography experiments. Its exact molecular mass is 61.08 Da. The selected N2C binders against caspase-3 have an average molecular mass of 14.8 kD and elute with an apparent molecular mass of 15 kD. The expected theoretical mass of the 2:1 complexes would be 90 kD or slightly higher. Caspase-3 is globular and heterodimeric, the overall shape of the complexes will be elongated, and consequently elute at a slightly shorter retention time. Therefore, the selected binders A2, A4, A12, C7, G5 and H2 in complex with caspase-3, displaying apparent molecular masses above 90 kD, are expected to be strong binders, whereas A7 and C8 with apparent molecular masses around 80 kD are expected to be weaker binders. The peak corresponding to the N2C ankyrin disappears upon complex formation (Figure 2_A), which is a good visual indicator for strong binders.

Two binders precipitated during incubation and could not be used for further studies (Table 1_A). An example of the elution profile of caspase-3 in complex with the selected N2C ankyrin binder A12, and the elution profiles of caspase-3 and A12 is shown in Figure 2_A.

Table 1_A: Selected N2C binders in complex with caspase-3, analyzed on a S200 analytical size exclusion column

Selected N2C binder in complex with caspase-3	Theoretical mol. mass	comment
A2_complex	101 kD	Strong 2:1 complex
A4_complex	96 kD	Strong 2:1 complex
A7_complex	81 kD	Weak complex
A12_complex	93 kD	Strong 2:1 complex
C7_complex	101 kD	Strong 2:1 complex
C8_complex	79 kD	Weak complex
D7_complex	X	Precipitation
G2_complex	X	Precipitation
G5_complex	97 kD	Strong 2:1 complex
H2_complex	95 kD	Strong 2:1 complex

SDS-PAGE analysis[3]

Fractions from analytical size exclusion experiments were analyzed on SDS-PAGE. Figure 3_A shows the selected binders A4 (1) and H2 (2) in complex with caspase-3 separated on a 15% SDS-PAGE gel. The three bands correspond to the α- and β-subunit of caspase-3 (18.6 kD and 11.8 kD, respectively) as well as the N2C binder (14.8 kD).

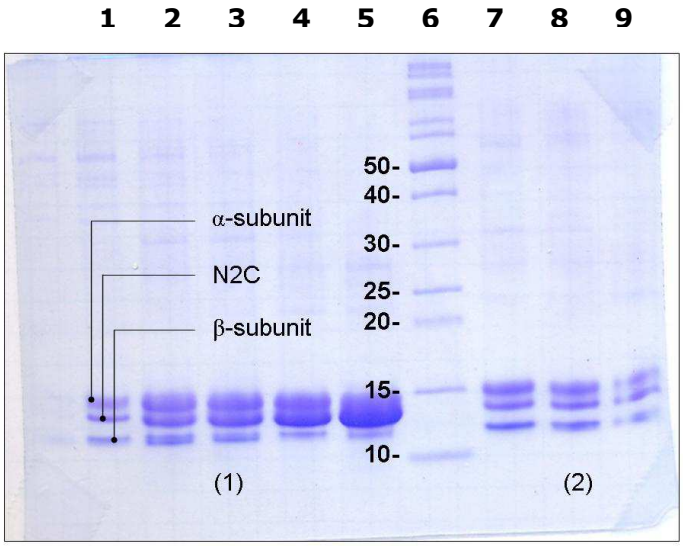


Figure 3_A)
A4_N2C (1) and H2_N2C (2) co-eluted with caspase-3, separated on a SDS-PAGE. Lanes 1-5: fractions (0.5 ml) of the complex peak of A4 and caspase-3; lane 6: marker; lanes 7-9 fractions (0.5 ml) of the complex peak of H2 and caspase-3. 20 µl of each fraction were applied per slot. Marker bands are indicated.

[3] Sodium dodecyl sulfate polyacryl gel electrophoresis; SDS-PAGE

50% Inhibitory concentration (IC50)

The IC50 value indicates the concentration, which is required to limit the enzymatic activity to the half-maximal turnover rate. It does not represent a kinetic parameter and is not comparable to the dissociation constant (K_D). Increasing amounts of inhibitor were mixed in caspase assay buffer (see appendix) with a constant amount of caspase-3 (1 nM) and incubated at 4°C for 2 hours. After substrate addition (Ac-DEVD-AMC; acetyl-aspartyl- glutamyl- valinyl- aspartyl-amino-methyl-coumarin) AMC was detected at 405 nm wavelength (Figure 4_A).

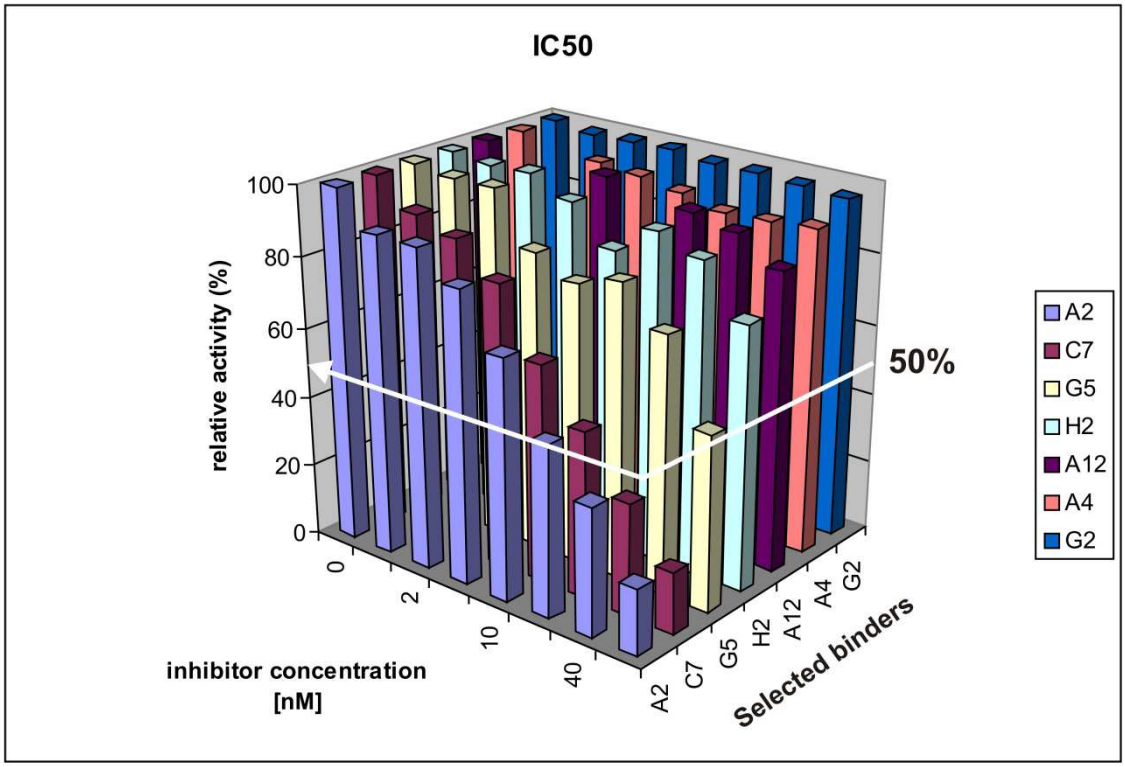


Figure 4_A)
Plot of inhibitor concentrations of selected N2C binders versus relative caspase-3 activity determined with the fluorogenic substrate Ac-DEVD-AMC at 405 nm wavelength on a plate reader (Genion, Tecan). The solid white line delineates the half-maximal turnover rate (50%).

The binders A2 and C7 show 50% inhibition of caspase-3 activity at a concentration of 20 nM, and the binder G5 decreases caspase-3 activity to 50% at 100 nM concentration. Other N2C binders identified by RD against caspase-3 were also tested. However, they showed either no, or rather weak (80%) inhibition at a 100 fold concentration excess (100 nM), and hence were not used for further studies.

APPENDIX PART II

Chapter 4

Construct sequence of NALP1 PYD with N-terminal GB1 tag, and C-terminal 6xHis tag:
MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGSMAGGAWG
RLACYLEFLKKEELKEFQLLLANKAHSRSSSGETPAQPEKTSGMEVASYLVAQYGEQRAWDLALHTW
EQMGLRSLCAQAQEGAGHHHHHHH

Construct sequence of TMS1/ASC PYD with N-terminal GB1 tag, and C-terminal 6xHis tag:
MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGSMGRARDAI
LDALENLTAEELKKFKLKLLSVPLREGYGRIPRGALLSMDALDLTDKLVSFYLETYGAELTANVLRDMG
LQEMAGQLQAATHQLEHHHHHH

(The first 56 aa represent the N-terminal GB1-tag followed by a GS linker, and the domain sequence) The constructs were provided by A. Kohl

Expression media for ^{15}N-labelled NALP1 PYD:

Vitamin mix (for 100ml)

- Thiamine	50 mg	
- d-Biotin	10 mg	
- Choline chloride	10 mg	
- Folic acid	10 mg	
- Niacin amide	10 mg	
- D-Panthothenic acid	10 mg	
- Pyridoxal	10 mg	
- Riboflavin	1 mg	(10 ml Vitamin mix / l M9)

M9 Medium /L culture

- ddH_2O	865 ml	
- M9 salts 10x	100 ml	to be autoclaved
- $MgSO_4$ 1M stock	2 ml	
- $NH_4Cl^{\#}$	1 g	
- Vitamine mix	10 ml	
- Solution Q	2 ml	
- Glucose 20%	20 ml	
- Ampicillin 100 mg/ml	1 ml (0.5, when in D_2O)	
- Chloramphenicol 25 mg/ml	1 ml (0.5, when in D_2O)	

$^{15}NH_4Cl$ for ^{15}N-labelled protein
all solutions above have to be sterile filtered before addition to M9 medium
($^{15}NH_4Cl$ can be added to M9 salts 10x)

M9 salts 10x (for 1 l culture)

- $Na_2HPO_4*2H_2O$	85.1 g
- KH_2PO_4	30 g
- NaCl	5 g

Alternatively $^{15}NH_4Cl$ added here!
pH adjusted to 7.2 (can be higher e.g. 7.4)

Solution Q (for 1 l culture; 1-2 ml / l M9)

- Conc. Hydrochloric acid (5M)	8 ml
- $FeCl_2*4H_2O$	5 g
- $CaCl_2*2H_2O$	184 mg
- H_3BO_3	64 mg
- $CoCl_2*6H_2O$	18 mg
- $CuCl_2*2H_2O$	4 mg
- $ZnCl_2$	340 mg
- $Na_2MoO_4*2H_2O^{\#}$	605 mg
- $MnCl_2*4H_2O$	40 mg

can be left away

Expression procedure for proteins in M9 minimal medium

Cells: BL21 pLys // Construct: GB1NalppyD long form c6xHis in pET20

5 ml LB medium supplied with appropriate amounts of ampicillin and chloramphenicol were inoculated with single colonies from a freshly transformed LB (better 2YT) agar plate and grown O/N at 37°C under vigorous shaking (250 rpm). These O/N cultures were centrifuged for 5` at 3000 g and resuspended in 2 ml LB. Intermediate cultures (50 ml LB, amp / cam) were inoculated with 1 ml resuspended cells, growth for 3h until OD_{600} ~ 0.7 is reached. In order to inoculate 250 ml LB (2 L flasks) with equal amounts of cells, ratios were calculated according to cell densities in the intermediate cultures. Further growth at 37°C until OD_{600} ~ 0.7 is reached. In the meantime, M9 media (100 ml aliquots, 250 ml flasks) was prewarmed. The cells were centrifuged and washed in M9 (without NH_4Cl). Finally, the cells were resuspended in M9 supplied with vitamin mix, antibiotics, glucose and magnesium sulphate. Protein expression was induced with 1 mM IPTG for 14 h at 18°C with moderate shaking.

Protein expression and purification

The PYD domain of NALP1 and the PYD domain of TMS1/ASC were cloned into a modified pET20 vector (Novagen), which produced the recombinant protein with an N-terminal GB1 solubility enhancer tag (B1 domain of Protein G) and a C-terminal hexa-histidine (His_6) sequence. The PYD domain of ASC was also cloned into a standard pET20 vector (Novagen) which produced the recombinant protein with an N-terminal hexa-histidine (His_6) sequence. Almost uniformly ^{15}N-labelled GB1_NALP1-PYD was obtained overexpressing the protein in BL21 pLys cells grown in 100 ml minimal medium (M9) containing $^{15}NH_4Cl$ as the sole heavy nitrogen source, vitamin mix, solution Q (see Appendix I), antibiotics (100 µg/ml ampicillin, 25 µg/ml chloramphenicol), 0.5% glucose and 1 mM $MgSO_4$. Cells were grown at 37°C to an optical density of OD_{600} = 1.5 – 2 cooled down to 18°C and then induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 14 hours O/N. The cells were resuspended in 50 mM Tris, 100 mM NaCl, 3 mM β-mercaptoethanol, 2 mM PMSF, 1 mM CHAPS, 1 mM imidazole (pH 8.0 at RT), lysed with a French press and centrifuged. The protein was isolated by Ni^{2+} affinity chromatography and subsequently purified by size exclusion chromatography on a S75 26/60 column (Amersham, Pharmacia Biotech) thereby rebuffering in 20 mM Tris, 100 mM NaCl, 0.2 mM CHAPS, 0.1 mM EDTA, 10 mM DTT pH 7.5 at 4°C. This buffer was used to measure interactions using NMR spectroscopy.

Unlabelled GB1_TMS1/ASC-PYD and ASC-PYD was obtained overexpressing the protein in BL21 pLys cells grown in 500 ml LB cultures supplied with ampicillin (100 µg/ml) and chloramphenicol (25 µg/ml). Cells were grown at 37°C to an optical density of OD_{600} = 0.6 – 0.8 cooled down to 18°C and then induced with 1 mM IPTG for 14 hours O/N). The cells were resuspended in 50 mM Tris, 100 mM NaCl, 2 mM PMSF, 0.5 mM CHAPS, 1 mM imidazole (pH 8.0 at RT), lysed with a French press and centrifuged. The protein was isolated by Ni^{2+} affinity chromatography and subsequently purified by size exclusion chromatography on a S75 26/60 column (Amersham, Pharmacia Biotech) thereby rebuffering in 20 mM Tris, 500 mM NaCl, 0.2 mM CHAPS, 0.1 mM EDTA, 10 mM DTT pH 7.5 at 4°C. This buffer was used to measure interactions using NMR spectroscopy.

Mass determination
The QTOF MS ESI was performed at the Functional Genomics Centre Zürich (FGCZ)

Dynamic light scattering
Dynamic light scattering experiments of both GB1 tagged domains were performed on a DynaPro Titan spectrometer (Wyatt technology corp., Charlottesville, USA) (data not shown) in a buffer containing 20 mM Tris, 100 mM NaCl, 0.2 mM CHAPS, 0.1 mM EDTA, and 10 mM DTT pH 7.5 at 4°C. The protein concentration was 1 – 2 mg/ml.

Circular dichroism
Circular dichroism spectra were recorded on a Jasco J-715 instrument (Jasco, Japan) at 20°C with a wavelength scan from 250 nm to 180 nm, data pitch of 1 nm, 4 seconds response time and a band width of 1 nm. The buffer contained 20 mM NaH2PO4, pH 4 (RT) with a protein concentration of 0.3 mg/ml or 20 mM Tris, 100 mM NaCl, 0.2 mM CHAPS pH 7.5 (RT) with a protein concentration of around 1.2 mg/ml.

APPENDIX PART III

Ribosome display and crude extract ELISA protocols

Constructs of caspase-3
Plasmid – caspase-3 active wild type // vector: pet11d // no tag // NcoI, BamHI sites

Caspase-3 Protein seq : C3P17-RG
MSGISLDNSYKMDYPEMGLCIIINNKNFHKSTGMTSRSGTDVDAANLRETFRNLKYEVRNKNDLTRE
EIVELMRDVSKEDHSKRSSFVCVLLSHGEEGIIFGTNGPVDLKKITNFFRGDRCRSLTGKPKLFIIQAC
RGTELDCGIETD

Caspase-3 Protein seq : C3P12-RG
MASGVDDDMACHKIPVEADFLYAYSTAPGYYSWRNSKDGSWFIQSLCAMLKQYADKLEFMHILTRV
NRKVATEFESFSFDATFHAKKQIPCIVSMLTKELYFYH

Caspase-3 DNA seq : C3P17-RG
ATG AGC GGT ATC TCC CTG GAC AAC AGT TAT AAA ATG GAT TAT CCT GAG ATG GGT
TTA TGT ATA ATA ATT AAT AAT AAG AAT TTT CAT AAA AGC ACT GGA ATG ACA TCT CGG
TCT GGT ACA GAT GTC GAT GCA GCA AAC CTC AGG GAA ACA TTC AGA AAC TTG AAA
TAT GAA GTC AGG AAT AAA AAT GAT CTT ACA CGT GAA GAA ATT GTG GAA TTG ATG
CGT GAT GTT TCT AAA GAA GAT CAC AGC AAA AGG AGC AGT TTT GTT TGT GTG CTT
CTG AGC CAT GGT GAA GAA GGA ATA ATT TTT GGA ACA AAT GGA CCT GTT GAC CTG
AAA AAA ATA ACA AAC TTT TTC AGA GGG GAT CGT TGT AGA AGT CTA ACT GGA AAA
CCC AAA CTT TTC ATT ATT CAG GCC TGC CGT GGT ACA GAA CTG GAC TGT GGC ATT
GAG ACA GAC TAA

Caspase-3 DNA seq : C3P12-RG
ATG GCT AGT GGT GTT GAT GAT GAC ATG GCG TGT CAT AAA ATA CCA GTG GAG GCC
GAC TTC TTG TAT GCA TAC TCC ACA GCA CCT GGT TAT TAT TCT TGG CGA AAT TCA AAG
GAT GGC TCC TGG TTC ATC CAG TCG CTT TGT GCC ATG CTG AAA CAG TAT GCC GAC
AAG CTT GAA TTT ATG CAC ATT CTT ACC CGG GTT AAC CGA AAG GTG GCA ACA GAA
TTT GAG TCC TTT TCC TTT GAC GCT ACT TTT CAT GCA AAG AAA CAG ATT CCA TGT ATT
GTT TCC ATG CTC ACA AAA GAA CTC TAT TTT TAT CAC TAA

(the vectors containing the p12 and p17 fragments of caspase-3 were obtained form R. Ganesan)

Caspase-3 purification protocol

The protocol is modified for the purification of caspase-3 based on this primary literature:

- Garcia-Calvo et al, Purification and catalytic properties of human caspase family members, 1999, 6(4), 362-9
- Stennicke, HR and Salvesen, GS (2000) Caspase assays. Methods Enzymol 322, 91–100

Human recombinant caspase-3 was produced in *E. coli* as inclusion bodies, refolded and purified according to the following protocol, obtained from R. Ganesan: Both subunits were separately expressed in BL21 CodonPlus (DE3) RIL cells. The cultures were grown to a density of A_{600} = 0.5 at 37°C in 0.5 L LB-medium (2 L flasks, vigorous shaking ~110 rpm) and expression was induced by the addition of IPTG (1 mM). The culture was shaken at 37°C for 4 hours post induction. The cells were harvested by centrifugation at 3000 g for 15 min, resuspended in 25 mL PBS buffer (8 g NaCl, 0.2 g KCl, 1.45 g Na_2HPO_4, 0.15 g KH_2PO_4 pH 7.4 in 100 mL) and lysed using a French Pressure cell. Since protein is predominantly present in inclusion body (IB) portion, the sample after cell lysis was centrifuged for 60 min at 20,000 g and the supernatant was discarded. The pellets were re-suspended in IB-wash-buffer A (25 mM HEPES pH 7.5, 1% CHAPS, 1 mM EDTA) using a homogenizer. After centrifugation for 30 min at 20,000 g the pellets were resuspended in IB-wash-buffer B (25 mM HEPES pH 7.5, 1 M urea) using a homogenizer. At least 3 cycles of washing with IB-wash-buffer-A & B were performed. The IB obtained after 3 washing steps were dissolved overnight at room temperature by slow rotation, in a buffer containing 6.5 M guanidine hydrochloride, 25 mM Tris pH 7.5, 5 mM EDTA, 100 mM DTT (Volume app. 10 mL per L culture).

Next day, the samples were centrifuged for 30 min at 20,000 g to remove any undissolved substances. This denatured protein sample can be aliquoted and stored at -80°C. Refolding was achieved at room temperature by rapid dilution to a final concentration of about 100 ug of subunit/L in refolding buffer (100 mM HEPES, pH 7.5, 10% sucrose, 1% CHAPS, 100 mM NaCl and 10 mM DTT). The p12 subunit was premixed with the p17 subunit in a 2:1 ratio. The solution was then transferred into an AMICON concentrating cell with a membrane cut-off of 10 kDa at 4°C. In order to reduce the salt concentration and to facilitate binding to an anion exchange column, the refolded protein was up-concentrated to 10-20 ml, diluted the first time with 20 mM Tris, 10 mM DTT, and the second time with anion exchange buffer A (20 mM Tris, pH 8.0, 20 mM NaCl and 10 mM DTT). 50 ml were loaded using a sample pump onto a 1 ml ResQ column (Pharmacia Biotech). The protein was eluted in a continuous gradient with anion exchange buffer B (20 mM Tris, 500 mM NaCl, 10 mM DTT, pH 8.0). The peak fractions were collected, concentrated with centricon, and loaded (500 uL) on a S200 (23 ml) size exclusion chromatography column equilibrated with 20 mM Tris, 150 mM NaCl pH 8.0. Fractions containing pure and active caspase were pooled and concentrated by ultra filtration (centricons-10,000 Da cut-off) to a final concentration of 10 mg/mL and the final yield was about 3- 4 mg/L culture.

Caspase assay

The caspase activity was determined from the initial rate of hydrolysis of substrate by measuring the accumulation of the fluorogenic product 7-amino-4-methylcoumarin. The increase in sample fluorescence was measured at an excitation wavelength of 360 nm and an emission wavelength of 465 nm for 5 min at room temperature (22-25°C) using an HTS 7000 Plus Bio Assay plate reader (Perkin Elmer). Both, the enzymes and the substrates were diluted in the assay buffer consisting of 20 mM piperazine-N,N`-bis (2-ethanesulfonic acid) (PIPES) pH 7.2, 100 mM NaCl, 10 mM 1,4-dithiothreitol (DTT), 1 mM EDTA, 0.1% (w/v) 3-((3-cholamidopropyl)-dimethyl-ammonio)-1-propane sulfonate (CHAPS) and 10% (w/v) sucrose. The typical assay volume was 150 ul in a Greiner 96-well black, flat bottom microtitre plate.

Ribosome display Selection Protocol

1. Coating plates

1.1 Neutravidin plate coating (done the day before the RD round)

- Neutravidin stock: 1.2 mg/ml in TBS; 20 uM (dilution 1:300 in TBS final conc. 66 nM)
- Pipet 100 ul of a 66 nM Neutravidin-TBS solution into the wells (duplicates)
- Seal with sticky plastic tape
- Store overnight at 4°C (no shaking required)

1.2 BSA-blocking of neutravidin coated plates

- Wash overnight incubated plate 2x with 1xTBS (flask), beat dry on paper
- Add 300 ul 0.5% BSA in 1xTBS to the wells (or WBT-BSA)
- Seal with sticky plastic tape
- Incubate on Shaker for 1 h at RT (7-800 rpm)

1.3 Immobilization with antigen

- Wash plate 3x with 1xTBS (flask), beat dry
- Add 200 ul 0.5% BSA in WBT (WBT-BSA) to each well (depends on target)
- Add 10 ul Antigen of approx. 10 uM conc. to the respective wells. Use biotinylated Antigen (with AVI-tag) void of free biotin
- Seal with sticky plastic tape
- Incubate on shaker for 1 h at 4°C
- Thaw Extract SL119(100 ul aliquots, -80°C) and Premix SL119 (-20°C) after 30`

Preparation before next step:

30` after Antigen immobilization:
- Thaw Extract SL119(100 ul aliquots, -80°C) and Premix SL119 (-20°C)
- Prepare bottle with liquid nitrogen

15` before translation start
- thaw RNA, and Heparin
- label tubes with "Name and stop"

10` before Translation:
- Wash plate with 3x 200 ul WBT-BSA each, beat dry
- Incubate wells that will not directly be used for panning with 128 ul WBT-BSA (Identical amount of liquid as will be used for panning)

2. Translation

2.1 In vitro translation

Mix according to pipetting scheme:
- Extract (100 ul aliquot) SL119 (contains all ribosomes from cells)
- Premix SL119 (vortex quickly) (contains amino acids)
- → 6 setups for 5 samples (always one more) (10 wells = 6 setups)
- This mixture can shortly be kept on ice!
- When RNA is thawed: add the appropriate amount of H_2O to RNA (2.5 ug/ul), transfer 2 ul into new tube; immediately N_2 freeze all RNA containing tubes
- After dispensing 2 ul RNA into labelled tubes, quick freeze tube in N_2 and keep on ice
- Add the mix to 2 ul 2.5 ug/ul RNA (prepipetted and N_2-frozen) and immediately start translation for 8 min 30 sec at 37°C on the heat block

t_o: add 53 ul Premix/Extract to RNA, dissolve RNA pellet, put tube at 37°C
30`` add again 53 ul Premix/Extract to 2nd sample
1`: add again 53 ul Premix/Extract to 3rd sample
1`30``: "
2`: "
- During translation prepare stopping buffer (WBT-BSA plus 12.5 ul/ml Heparin à 200 mg/ml) in previously labelled RNase free tubes, keep on ice (see pipetting scheme)
8`30``: remove translation mix, flick shortly and transfer to stopping buffer
9`: remove 2nd sample
9`30``: remove 3rd sample
10`: "
10`30``: " and keep them on ice
Centrifuge all tubes at 20000 g for 5min at 4°C, keep tubes on ice

2.2 Stop reaction

(If not done before) during centrifugation wash all wells 3x with 300 ul WBT-BSA (transfer of Mg^{2+}!)
- Add 128 ul into "panning" or prepanning wells respectively
- keep tubes on ice and plate always at 4°C

3. RD in plates

The translated tertiary Ribosome-RNA-Protein complex first has to incubate in the "prepanning" well and is further transferred into the true panning wells.
Always perform 2 prepanning steps! (In duplicates)

3.1 Panning at temperatures equal or below 4°C)

- Pipet 128 ul stopped translation mix to the first column of prepanning wells (duplicate)
- Seal with sticky plastic tape and incubate for 45 min at 4°C in the shaker (prepanning)
- Remove 128 ul WBT-BSA from the 2nd prepanning wells
- Transfer the 128 ul prepanning mix into the 2nd prepanning wells, seal and incubate for another 45 min at 4°C in the shaker
- Finally transfer the supernatant into the panning wells, seal and incubate for another 45 min at 4°C in the shaker

Washing: General washing principle
- 3x wash - 1x incubate (30 min, 1 h, 2 h, 3 h) - 3x wash (totally 7x)
→ In the first round: remove lid and add 3x 200 ul WBT-BSA into the wells, beat dry
- Go ahead without long incubations
In later rounds: As in first, but make long washing incubations (30 min, 1 h, 3 h)

Preparation before RNA elution:
10-20` before end of incubation:
- Cool down a centrifuge to 4°C (tighten lid, select fast cool procedure)
- Set two heating blocks to 50°C and 70°C
- Prepare 3 series of tubes (duplicates! = 3 tubes for each well to be eluted and label them with sample/exp. "lysis", "eluted RNA", "RT", name and round number
- Add 400 ul lysis buffer of Roche RNA-purification kit into "lysis" tubes
- Prepare Roche-RNA purification columns for each sample and label them

Just before elution:
- thaw the reagents for RT (WTC4, DTT and buffer)

3.2 Elution (4°C room)

- Make sure that the RNA is kept at 4°C or below!
- Elution buffer: EB25 (+50 ug/ml *S.cerevisiae* RNA; 2 ul of a 25 ug/ul stock per ml)
- Prepare 2x 1.2 ml EB25 + 2.2ul RNA (5 targets = 10 wells = min. 2 ml)
- Add 100 ul elution buffer incubate on shaker for 5min and pipet the eluate directly into previously prepared chilled 400 ul Lysis buffer (Roche) and vortex quickly to mix well

- Repeat this elution step, again 5 min elution and vortex quickly
- The RNA is stable in the lysis buffer and can be brought to room temperature

4. RT-PCR I

4.1 RNA purification using Roche high pure RNA isolation kit

- Just before RNA-purification thaw the reagents for RT (WTC4, DTT and buffer)
- All centrifugation steps are performed at 4°C
- Apply the lysis buffer/eluate mixture onto the column
- Spin for 1 min at 8000 g, discard flow through
- Wash with 500 ul buffer 1 (black cap), discard flow through (manual short spin 8000 g)
- Wash with 500 ul buffer 2 (blue cap), discard flow through (manual short spin 8000 g)
- Add 200 ul buffer 2 (blue cap) and spin down for 2 min at 13000 rpm
- If required, wait with the RNA elution (should be stable)
- Add 30 ul elution buffer and incubate for 1 min keep tubes on ice, elute RNA (at 8000 g for 1 min and directly transfer the RNA to 70°C for 10 min.
- During RNA denaturation prepare Master mix (pipetting scheme) for RT-PCR (Inner primer WTC4) and spread mix into previously labelled tubes keep on ice!

4.2 Reverse Transcription (RNA → DNA)

- Immediately start RT-PCR right after 70°C incubation of the RNA
- Flick tubes, spin down the denatured eluted RNA and keep on ice
- Add 12.25 ul eluted RNA to RT mix and N_2 freeze the rest of the RNA
- Put at 50°C for 1h (can be less)
- Store this RT-mixture on ice
- Mastermix: Primer (WTC4)
    - DTT
    - Buffer
    - Enzymes

4.3 PCR on RT (DNA amplification of binding region) - 1st PCR

- Pipetting scheme: use INNER PRIMERS (EWT5/WTC4)
- Always take a fresh aliquot of VENT polymerase for this amplification, following PCR`s can be performed with the leftover; thaw it up just before PCR
- Prepare mastermix without VENT
- Program (to be adapted according to primers and template) name: RIBDIS (RD50)
- Hot start: start program and press pause; the program stops, but the lid heats up
- Prepare PCR tubes with 5ul RT-mixture and N_2 freeze the rest
- Start the program (whole block has to be at 95°C)
- Add VENT into mastermix just before PCR start and equally dispense 45 ul
- Include positive control with known DNA and negative control with DNA elution buffer

5. RT-PCR II

Program:

5`	95°C		
30``	95°C		
30``	55°C	25-45 cycles depending on RD round	
45``	72°C	45x – 1st	
		35x – 2nd	
5`	72°C	30x – 3rd	
	4°C	25x – 4th	

- Verify the PCR on a 1.2% agarose Gel (2 ul PCR product + 2 ul 2x loading buffer)
- If the agarose Gel is OK make a preparative Gel, cut the band and purify the Product with a QIAGEN Gel purification Kit

5.1 PCR on RT-PCR - 2nd PCR

Mix the following reaction according to pipetting scheme (Inner primers EWT5/ WTC4)
Program to be adapted

5`	95°C	
30``	95°C	
30``	55°C	25-45 cycles depending on RD round
45``	72°C	
5`	72°C	
	4°C	

- Verify the PCR on a 1.2% agarose Gel (2 ul PCR product + 2 ul 2x loading buffer)
- If the agarose Gel is OK make a preparative Gel, cut the band and purify the Product with a QIAGEN Gel purification Kit

6. Cloning into pRDV

6.1 Digest

DNA	5 ul (2^{nd} PCR step from 4.4)	
NcoI	1 ul	
HindIII	1 ul	
NEB2	2 ul	
H_2O	11 ul	60 min, 37°C

Use MinElute PCR purification kit for the digested fragments and elute with 10ul
(5.1.2 Digest pRDV (only done if plasmid stock is empty))

pRDV	x uL	
NcoI	3 uL	60 min, 37°C
HindIII	2 uL	+ 3 uL CIP
NEB2	5 ul	60 min, 37°C
H_2O	xx to 50 ul	

6.2 Ligation

pRDV	2 ul (check concentration) can be less e.g. 1.5 ul
PCR product	9 ul whole MinElute eluate
T4 DNA Ligase	1 ul
10x LB	1.5 ul
H2O	xx ul

Leave ligation mix for 60 min at RT on bench, then put ligation mix at 70°C for 10 min

7. PCR of pRDV - 3^{rd} PCR

Mix the following reaction according to pipetting scheme (OUTER PRIMERS T7B/TolAk)
Adapt program

5`	95°C	
30``	95°C	
30``	55°C	30-35 cycles depending on RD round
1`	72°C	
5`	72°C	keep for infinite time at 4°C

- Verify on small 1.2% agarose Gel
- Purify product via preparative 1.2% agarose gel (better than PCR purification Kit)

8. PCR for Transcription - 4th PCR

Mix the following reaction according to pipetting scheme (OUTER PRIMERS T7B/TolAk)
Same program as before:

5`	95°C	
30``	95°C	
30``	55°C	30-35 cycles depending on RD round
1`	72°C	
5`	72°C	keep for infinite time at 4°C

- Verify on a small 1.2% agarose Gel
- Do not purify this product!!!

9. Transcription of PCR products into mRNA

9.1 Transcription (50%)

- Mix the reaction according to protocol
- Transcription for 2-3 h at 37-38°C (better 3 h)
- Strange smell, like a freshly opened can of corn

9.2 RNA isolation

- Add 100 ul ice-cold UHP
- Add 200 ul 6 M LiCl
- Keep on ice for 30 min
- Spin 20`000 g, 4°C, 30 min
- Remove supernatant carefully add 70% EtOH and remove it right afterwards
- Short spin
- Remove supernatant with small tips (200 ul pipette)
- Leave the pellet for 5`at RT (tubes remain open)
- Dissolve pellet in 200 ul ice-cold UHP (takes some time) vortex shortly
- Spin suspension 20`000 g 4°C, 5 min
- Transfer 180 ul supernatant into tubes containing 20 ul 3 M NaOAc
- Add 500 ul ice cold 100% EtOH, vortex
- Put at -20°C for 30`(or leave O/N at -20°C)
- Spin 20`000 g, 4°C, 30 min
- Discard supernatant
- Wash with 500 ul ice-cold 70% EtOH
- Short spin
- Discard supernatant
- Dry pellet in a speedvac apparatus

9.3 RNA aliquotation

- Take up the pellet in 30 ul ice cold UHP
- Immediately transfer 2 ul to 500 ul ice cold UHP for OD_{260} quantif. and N_2 freeze rest
- Measure OD_{260}
- For RNA, 1OD corresponds to a concentration of 40 ug/ml. At the here used dilution 1 OD corresponds to 10 ug/ul of stock RNA. Thus OD*10 = concentration in ug/ul
- Add UHP to the RNA before next RD round in order to reach an equal RNA concentration of all samples.

Crude Extract ELISA (Enzyme linked immunosorbent Assay with DARPins)

(Adapted protocol from Molecular Partners PL00003_CrudeExtract_ELISA.doc Version 20070213)
Protocol modified by Tobias Merz July 2007-07-06
(Minor modifications include indications of required amounts per plate and cancellations of unused material and procedures)

Protocol 01: Immobilization (Coating of plates)

Equipment and Reagents:

- Nunc MaxiSorp 96-well plate (Nunc, no. 442404)
- 25% Tween 20
- TBS (50 mM TrisHCl pH 7.4, 150 mM NaCl)
- 300 x Neutravidin (Neu) stock (Pierce, no. 31000) (1.2 mg/ml 20uM) Neutravidin dissolved in UHP, filter through 0.22 um filter, store at -80°C)
- TBS-T (TBS containing 0.1% (v/v) Tween20 (4 ml/l of 24% Tween20))
- TBS-TB (TBS-T containing 0.5% (w/v) BSA)
- Shaker for Nunc MaxiSorp 96-well plates

Method

Day1

- Coat a Nunc MaxiSorp 96-well plate with 100 ul per well of 66 nM Neutravidin (4 ug/ml, 1:300 dilution of stock solution) in TBS overnight at 4°C.

>> Required amount (96 samples) 2 x 96-well = 20 ml TBS, 66 ul Neutravidin stock

- Start with the expression of DARPins in the 96-well format

Day2 (in parallel with the expression of DARPins in the 96-well format)

Beat dry the plate on stacked paper towels. Add 300 ul/well of TBS-B blocking buffer and incubate the plate at room temperature (RT) with orbital shaking (900 rpm) for 1 hour.

>> Required amount (96 samples) 2 x 30ml TBS, 2 x 150 mg BSA

Beat dry the plate on stacked paper towels, add 300 ul/well TBS wash buffer. Repeat wash cycle 3 times.

Beat dry the plate on stacked paper towels. Add 100 ul/well TBS-B with 10-500 nM antigen (the higher the better) and incubate the plate at 4°C for 1 hour.

>> Required amount (96 samples) 1 x 96-well = 10 ml TBS-B + antigen

Wash the plate as in step 2.

Proceed to the ELISA: Protocol 03, Step 1

Every target needs a direct MBP background sample, thus 2 plates with 2 x 96-wells!

Protocol 02: Expression of DARPins in the 96-well format

Equipment and Reagents

- 96-deep well plates for bacterial cell cultures (e.g. AbGene, no. AB-0932)
- Isopropyl-β-D-thiogalactoside (IPTG) stock (1 M IPTG 238 mg/ml)
- Ampicillin (1000x stock) (50 mg/ml ampicillin 0.22 um filtered)
- 2YT media (1% bacto-yeast extract (10g/l), 1.6% bacto-tryptone (16 g/l), 0.5% NaCl (5 g/l)

LB media (premix, 25 g/l) or (0.5% bacto-yeast extract (5 g/l), 1% bacto-tryptone (10 g/l), 1% NaCl (10 g/l)

- 20% glucose (200 g/l) in UHP, autoclave and cool)

Method (performed with 8 fold multichannel pipette)

Day 1

- Aliquot 1.2 ml LB/Glu/Amp into each well of a 96-deep well plate (master plate). 2YT gives higher densities than LB and higher plasmid DNA yields.

>> required amount (96 samples) 123.5 ml LB, 6.5 ml 20% Glc, 130 ul Amp (1000x)

- Pick individual colonies into each well and seal the plate by applying an air permeable membrane. A simple method that avoids cross contamination and double inoculation is to seal the plate with a piece of parafilm during inoculation. Just stab the picked colonies through the parafilm into the medium. As negative control, leave at least one well without inoculation.
- Incubate sealed plate in an orbital shaker (450 rpm) and grow O/N at 37°C

Day 2 (Parallel continue preparation of Nunc MaxiSorp 96-well plates for the ELISA)

- Aliquot 0.9 ml of LB/Amp into each well of a 96-deep well plate (expression plate).

>> required amount (96 samples) 130 ml LB, 130 ul Amp (1000x)

- Transfer 0.3 ml from each well of the master plate into the corresponding well of the expression plate. (Estimated OD_{600} = 0.3, if properly grown)
- Incubate the sealed plate in an orbital shaker (450 rpm) for 1 hour, at 37°C
- Facultative: Spin master plate at 3000 g for 10 min, empty plate, blot dry store at -20° for later plasmid DNA preparations
- Add 50 ul LB/Amp/IPTG (0.5 mM) and incubate sealed plate at 37°C for 3-4 h at 450 rpm.
>> required amount (96 samples) 14 ml LB, 14 ul Amp, 175 ul IPTG (= 12.5 mM)
- Spin expression plate at 3000 g for 6 min, blot dry on paper towels
- Add 50 ul B-Per II to each well and resuspend pellet by vortexing 2 min, 10 – 15 min shaking
- Add 950 ul TBS 400 and spin plate at 3000 g for 10 min to remove cell debris.
- Proceed with the screening of DARPins as described in Protocol 03, Step 1
- For further experiments store the plate at -20°C

Protocol 03: Screening DARPins by ELISA

Comment

Include appropriate positive controls (MBP) and negative controls for ELISA

Equipment and reagents

- Antigen coated plate and with control lanes for background binding (Protocol 01)
- Anti-RGS(His)4 mAb (mouse anti RGS(His)4 IgG1 (Qiagen, no. 34650)
- Anti-mouse mAb AP conjugate (goat anti-mouse IgG (Sigma, no. A3562)
- pNPP substrate (3 mM (di-sodium 4-nitrophenyl phosphate (pNPP) (1.1 mg/ml), 50 mM $NaHCO_3$ (4.2 mg/ml), 50 mM $MgCl_2$ $6H_2O$ (10.2 mg/ml), store at -20°C) 23.1 mg pNPP, 88.2 mg $NaHCO_3$, 214.2 mg $MgCl_2$ $6H_2O$ in H2O (for 96 samples)
- 96-well plate containing crude extracts (Protocol 02) and a shaker for Nunc MaxiSorp 96-well plates

Method

- Transfer 100 ul of the supernatant form Protocol 02 onto the ELISA plate form Protocol 01 Step 05. Supernatants can be diluted up to 200 fold for further competition ELISAS.
- Incubate for 30-40 min at 4°C on orbital shaker (900 rpm).
- Empty plate and blot dry, wash 3 times all wells with TBS-T (300 ul/well approx.).
- add 100 ul 1:2000 Anti-RGS(His)4 mAb (1^{st} Ab) in TBS-TB and incubate for 30-45 min at RT.
>> required amount (96 samples) 21 ml TBS-T, 105 mg BSA, 10.5 ul Anti-RGS(His)4
- Wash plate as in step 3 (3 wash cycles).
- Add 100 ul 1:10000 Anti-mouse AP conjugate (2^{nd} Ab) in TBS-TB, incubate for 30-45 min at RT.
>> required amount (96 samples) 21 ml TBS-T, 105 mg BSA, 2.1 ul Anti- mAb-APconj.
Wash plate as in step 3 (4 wash cycles).
- Add 100 ul pNPP substrate solution for the detection; prepare 21 ml (for 96 samples).
- The reaction can be stopped by adding 3 M NaOH.
- Measure OD_{405} – OD_{450} in 15 min intervals depending on the color development.
- Perform further SDS-PAGE analysis of single clones if additional information is needed.
- Take 10 ul from Protocol 02 step 11, add 3 ul SDS gel loading buffer and heat the samples. Load 5-10 ul on SDS-PAGE
- Prepare plasmid DNA from positive clones using a standard kit (Qiagen) by directly adding resuspension buffer to the respective well of the master plate (thaw plate) by shaking (900 rpm) for 1 min.
- Strike out (still frozen) cells from the wells with a toothpick on a new plate, grow O/N, cultivate 3 ml O/N culture and prepare Mini/Midiprep to get higher plasmid DNA yields.

ACKNOWLEDGEMENTS

First, I would like to thank Markus Grütter, for his support and for allowing me to work in several different and highly interesting research fields. I was given the opportunity not only to investigate group inherent related apoptosis topics, but also to be involved in several crystallographic collaborations with the Polyphor Company, the Plückthun group and the Kohler group.

I would also like to thank various members of the biochemistry institute such as Prof. Dr. Raimund Dutzler and Prof. Dr. Andreas Plückthun for accepting the responsibility being my thesis co-referees, Prof. Dr. Antonio Baici, who always had an open door to discuss problems about enzyme kinetics and Dr. Sergio Gloor for giving me helpful advise how to teach and deal with students.

Then, I would like to thank Dr. Peer Mittl for stimulating discussions about crystallographic topics and his helpful comments and contributions for the two manuscripts written during my PhD thesis and Dr. Christophe Briand who always had a helping hand to solve computer problems, for explaining me computer software and to support me during many hours of data collection.

Furthermore, I owe many thanks to Dr. Andreas Kohl who coached me during my first year PhD study and showed me how to work scientifically and efficiently, and Dr. Daniel Frey who always took his time to explain and show me the tricks to make someone's life easier.

My sincere thanks go to Prof. Dr. Dieter Seebach and the members of the Kohler Group, in particular to Dr. Tobias Heck, who was a great help and support during the BapA project.

Finally, I would like to thank my family and colleagues, who found consoling and encouraging words during my thesis.

Ein Gedanke zum Thema Forschung:

… am Ende meines PhD Studiums komme ich zum Schluss, dass

Je mehr ich weiss und lerne,
desto dümmer scheint es mir,
das behaupt ich gut und gerne,
ich mich fühl' als Brevetier;

Paradoxerweis ist der der fragt,
und selbst Nichts lernt im täglich Tun,
nur scheinbar klüger, wenn er sagt:"
genau so hätt' ichs auch gemacht".

CURRICULUM VITAE

Persönliche Angaben

Name	Merz
Vorname	Tobias
Geburtsdatum	10.02.1977
Geburtsort	Olten (SO)
Heimatort	Beinwil a. See / Leimbach

Ausbildung

Alte Kantonsschule Aarau, Typus E 1993 – 1997

(Wirtschaftliches Gymnasium)

Militär 1997 - 1998

Rekrutenschule, Isone; UOS, Herisau

Biochemiestudium an der Universität Zürich 1998 - 2003

Diplomarbeit (Master) Okt. 2002

Cloning, Expression and Purification of the Cadherin-Like Domains of Human Calsynthenin-1 asGST Fusion Proteins

Unter der Leitung von Prof. Dr. P. Sonderegger

Diplomprüfung Jan. 2003

PhD, Biochemisches Institut Universität Zürich 2003 - 2010

Crystallographic and Biochemical Analysis of a Full Consensus Designed Ankyrin Repeat Protein, of Proteins Involved in Human Host Defence and a Beta-peptide Hydrolase

Unter der Leitung von Prof. Dr. Markus G. Grütter